

Computational Optimization of Metal-Organic Framework (MOF) Arrays for Chemical Sensing

by

Jenna Ann Gustafson

Bachelor of Science, Chemical Engineering, University at Buffalo, 2014

Submitted to the Graduate Faculty of
the Swanson School of Engineering in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH
SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Jenna Ann Gustafson

It was defended on

February 19, 2019

and approved by

Christopher E. Wilmer, Ph.D., Assistant Professor
Department of Chemical and Petroleum Engineering

Eric Beckman, Ph.D., Distinguished Service Professor
Department of Chemical and Petroleum Engineering

Susan Fullerton, Ph.D., Assistant Professor
Department of Chemical and Petroleum Engineering

Paul Ohodnicki, Ph.D., Lecturer
Department of Electrical and Computer Engineering

Dissertation Director: Christopher E. Wilmer, Ph.D., Assistant Professor
Department of Chemical and Petroleum Engineering

Copyright © by Jenna Ann Gustafson

2019

Computational Optimization of Metal-Organic Framework (MOF) Arrays for Chemical Sensing

Jenna Gustafson, PhD

University of Pittsburgh, 2019

Although commercial gas sensors exist for applications such as product quality control, industrial food monitoring, and smoke detection, there are many potential applications for which adequate gas sensing technology is lacking. There is an unmet need for gas sensors to detect natural gas leaks, for disease detection via breath analysis, and for environmental monitoring, to name just a few examples. Current gas sensors do not exhibit the sensitivity and/or selectivity required to detect trace amounts of the required gases in complex gas mixture environments (e.g., ambient air or a patient's breath). It is known that arrays of sensors, or electronic noses, improve chemical detection when compared to single sensor elements. Although some work has been done to optimize sensor device performance, there are many potential sensing materials that have not yet been extensively explored.

Herein, we explore the use of metal-organic framework (MOF) materials in sensor arrays, exploiting their high adsorption capabilities to yield more selective and sensitive electronic noses. As a relatively new class of materials, MOFs have not been thoroughly investigated for gas sensing applications. In particular, prior to our work, there had only been a few investigations of MOF sensor arrays and those were limited to purely experimental work that relied heavily on trial-and-error. We demonstrate that leveraging computational modeling and optimization to rationally design MOF sensor arrays can yield significantly improved sensing performance.

Our novel computational method was carried out first by predicting individual MOF sensor responses via molecular simulations. Then, we developed a method to analyze those individual responses and provide output signals for entire sensor arrays to predict unknown gas mixtures. Following this, the prediction ability of each array was evaluated according to the Kullback-Liebler divergence (KLD), where we determined the best arrays for detecting methane-in-air mixtures. Finally, we developed and validated a genetic algorithm that enables the optimization of large MOF arrays.

Table of Contents

Preface.....	xiv
1.0 A Brief History of Gas Sensing Technologies: Motivation for New Approaches and Improved Sensing Materials to Optimize Electronic Nose Design.....	1
1.1 An Introduction to Gas Sensors and Their Applications.....	1
1.2 Electronic Noses	6
1.3 New Classes of Sensing Materials	8
1.3.1 Carbon Nanotubes and Zeolites.....	8
1.3.2 Metal-Organic Frameworks.....	9
1.4 Computational Methods for the Analysis of Electronic Nose Signals	12
2.0 Computational Design of Metal-Organic Framework Arrays for Gas Sensing: Influence of Array Size and Composition on Sensor Performance.....	14
2.1 Molecular Simulations of Gas Adsorption	14
2.2 MOF Array Scoring Metric.....	18
2.3 Ranking MOF Arrays	19
2.4 Effect of Structure-Property Relationships on Sensing	24
2.5 Conclusions	25
3.0 Optimizing Information Content in MOF Sensor Arrays for Analyzing Methane-Air Mixtures	28
3.1 Background and Motivation.....	28
3.2 Performing Molecular Simulations of Gas Adsorption	31
3.3 Choosing “Experimental” Gas Mixtures.....	33
3.4 Simulating Sensor Element Mass Measurements.....	33

3.5 Combining Information from Sensor Elements	34
3.6 Ranking Sensor Arrays.....	35
3.7 Conclusion.....	46
4.0 Intelligent Selection of Metal-Organic Framework Arrays for Methane Sensing via Genetic Algorithm.....	48
4.1 Motivation	48
4.2 Genetic Algorithm	50
4.3 Predicting Gas Adsorption in MOFs	55
4.4 MOF Array Sensing	56
4.5 Genetic Algorithm Analysis.....	57
4.5.1 Comparison to Brute Force Analysis	57
4.6 Genetic Algorithm Parameter Analysis.....	60
4.6.1 Varying Population Size	60
4.6.2 Varying Mutation Strength.....	63
4.6.3 Increasing Array Sizes.....	65
4.7 Conclusions	67
5.0 Development of Metal-Organic Framework (MOF) Coated Bulk Acoustic Sensors for Carbon Dioxide and Methane Detection	68
5.1 Introduction and Motivation	68
5.2 Methodology.....	70
5.2.1 Experimental	70
5.2.2 Simulations	71
5.3 Results and Discussion	73
5.4 Conclusion.....	75

6.0 The Future of MOF-Based Electronic Nose Development.....	76
6.1 High Throughput Screening of MOFs for Sensing	76
6.1.1 Using physical Parameters as Filtering Criteria	76
6.1.2 Predicting Complex Gas Mixture Adsorption.....	77
6.1.3 Interpolation to Predict Gas Mixture Adsorption	78
6.1.4 Tiered Approach	79
6.2 Alternative Signal Transduction Mechanisms.....	80
6.2.1 Electrical Conductivity	80
6.2.2 Optical	81
6.3 Electronic Nose Architecture.....	81
6.4 Post-Simulation Data Analysis Techniques.....	83
Appendix A Computing Gas Mixture Adsorption.....	84
Appendix B Lists of Simulated Gas Mixtures	91
B.1 Sensor Array Gas Space Score	91
B.2 Kullback-Liebler Divergence Calculation	92
B.3 Genetic Algorithm Analysis	92
Appendix C Sensor Array Gas Space Score.....	94
C.1 Example Calculation	94
C.2 List of All Results.....	98
Appendix D Kullback-Liebler Divergence	99
D.1 Detailed Explanation of Algorithm	99
D.2 Example Calculation	102
Appendix E Genetic Algorithm	109

E.1 Input Files and Parameters.....	109
E.2 Calculation of Fitness Function	111
E.3 GA Analysis	117
E.4 Theoretical Limits of the Kullback-Liebler Divergence	119
Bibliography	121

List of Tables

Table 1. List of binary experiments tested with mole fractions of each gas.....	38
Table 2. List of ternary experiments tested with mole fractions of each component.	38
Table 3. All gas mixtures tested in this study, in mole fractions.	71
Table 4. Physical properties of MOF structures	84
Table 5. LJ Parameters for Gas Molecules	86
Table 6. LJ parameters for framework atoms in all MOFs.....	87
Table 7. Partial charges for gas molecule atoms.....	88
Table 8. Partial charges of framework atoms	88
Table 9. Atom positions for 3 dimensional molecules	89
Table 10. Critical constants for gas molecules, as seen in the PR EoS	90
Table 11. List of all gas mixtures simulated for SAGS score study, in mole fractions.....	91
Table 12. List of all gas mixtures simulated for information content optimization study, in mole fractions.....	92
Table 13. List of all gas mixtures simulated for SAGS score study, in mole fractions.....	93
Table 14. Three gas mixture compositions and total masses adsorbed in three different MOFs .	94
Table 15. List of All MOF combinations and their scores, ranging from 1-5 MOF arrays. Key: IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), ZIF-8 (Z)	98
Table 16. Sample list of gas mixtures and simulation results.....	102
Table 17. Experiment 1 composition and simulation results.....	102
Table 18. "Unknown" mixture 1 composition and simulation results	111

List of Figures

Figure 1. SAW device.....	4
Figure 2. MOFs as sensing materials on SAW sensor array.....	11
Figure 3. Adsorption (mg/g MOF framework) of CO ₂ (green triangles), ⁹² C ₂ H ₆ (blue circles), ⁷⁶ N ₂ (red squares), ⁹² and CH ₄ (orange diamonds) ⁹³ in HKUST-1 at 298K. Solid lines with filled symbols indicate experimental measurements taken from the literature and dashed lines with open symbols indicate our simulation predictions.....	16
Figure 4. (a) The five MOFs we studied as sensing materials that could be (b) layered/deposited onto a surface acoustic wave (SAW) sensor. (c) Devices made from arrays of multiple SAW sensors on which different MOF structures are layered: 5 x 1 MOF arrays, 10 x 2 MOF arrays, 10 x 3 MOF arrays, 5 x 4 MOF arrays, 1 x 5 MOF array (31 arrays in total).	17
Figure 5. <i>Sensor array gas space scores</i> for MOF arrays consisting of different combinations of one to five MOFs in an array, using IRMOF-1, HKUST-1, NU-125, UiO-66, and ZIF-8 at 298K and 1 bar.....	20
Figure 6. <i>Sensor array gas space scores</i> at 298K, 1 bar and 10 bar, for arrays with three elements using combinations of IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), and ZIF-8 (Z). The data at 10 bar is ordered by increasing score along the x-axis, but the corresponding scores for sensors at 1 bar do not follow the same trend.	21
Figure 7. (a) Percent increase in SAGS score as one MOF (specified by color) is added to an existing two-MOF array, specified on the horizontal axis, at 1 bar: IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), ZIF-8 (Z). (b) Percent increase in SAGS score as one MOF is added to an existing two-MOF array at 10 bar.....	22
Figure 8. (a) Percent increase in SAGS score as one MOF (specified by color) is added to an existing three-MOF array, specified on the horizontal axis, at 1 bar: IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), ZIF-8 (Z). (b) Percent increase in SAGS score as one MOF is added to an existing three-MOF array at 10 bar.....	23
Figure 9. (a) SAGS score of each MOF array versus the void fraction averaged over all MOFs in the array, at 1 bar (red) and 10 bar (blue) (b) SAGS score versus the surface area averaged over all MOFs in the array, in m ² /g (c) SAGS score versus the density averaged over all MOFs in the array, in g/cm ³ . Please see Table 4 in the Supporting Information for a detailed list of MOF properties.....	24
Figure 10. Flow diagram of analysis described in the Methodology section.	30

Figure 11. Unit cells of all MOFs used in this study: a) NU-100, b) HKUST-1, c) IRMOF-1, d) MgMOF-74, e) MOF-801, f) ZIF-8, g) UiO-66, h) MOF-177, and i) NU-125.....	32
Figure 12. Total mass adsorbed (normalized by subtracting a value from each data set so all start at 5 mg/cm ³ for a CH ₄ concentration of 0) for nine MOFs vs concentration of CH ₄ in various CH ₄ /N ₂ mixtures.....	36
Figure 13. Total mass adsorption (mg/cm ³) vs. mole fraction of various ternary gas mixtures in nine MOFs: a) IRMOF-1, b) HKUST-1, c) NU-125, d) UiO-66, e) ZIF-8, f) MgMOF-74, g) NU-100, h) MOF-801, i) MOF-177. Note: The color bar scales vary between plots. Color represents the total mass of gas adsorbed (mg/cm ³), and each triangle edge is the component mole fraction.	37
Figure 14. Average KLD vs number of MOFs in an array for the best (red) and worst (blue) arrays of each size over all four binary mixture experiments.....	39
Figure 15. Averages over nine experiments: a) KLD vs number of MOFs in an array, predicting CH ₄ (bar 1), N ₂ (bar 2), and O ₂ (bar 3). The best array KLD is read at the top of the red and the worst is the blue portion. b) This is the same as a) but averages the KLD over all three gases at each array size.	40
Figure 16. The probability versus concentration of CH ₄ , N ₂ , and O ₂ is plotted for the top and bottom two arrays, showing increasing confidence through narrowing peaks as the array size increases. Arrays are ranked by the product of the KLD values as they relate to each gas.	42
Figure 17. The nine-MOF probability vs concentration plot from Figure 16 above. Below are the five highest probability gas mixtures for experiment #1 in Table 2, whose concentrations align with the peaks in the plot.	43
Figure 18. Probability vs concentration for all simulated gas mixtures in the nine-MOF array, a) – i) being experiments 1–9.....	44
Figure 19. Probability vs concentration for all simulated gas mixtures for the best four-MOF arrays a) – i) being experiments 1–9.....	46
Figure 20. Possible combinations on a log scale vs array size, varying the possible number of materials to choose from (60, 50, 40, 30, 25).	49
Figure 21. Flow diagram of steps in a genetic algorithm.	50
Figure 22. Example of an encoded MOF array in a genetic algorithm.....	50
Figure 23. Procedure for creating the first population of arrays.....	51
Figure 24. Fitness (KLD) is calculated for each MOF array and ranked from highest to lowest. 52	
Figure 25. Selecting parents from the top 20% of ranked arrays and random 20% of arrays.	53

Figure 26. Crossover of two parents to create a child.....	53
Figure 27. Mutation of an individual, with a mutation rate of 0.1	54
Figure 28. KLD (in bits ³) vs ranking of every MOF array for size of a) 1 MOF, b) 2 MOFs, c) 3 MOFs, and d) 4 MOFs. Black points represent best arrays found from the GA over five runs, where some are overlapping.	58
Figure 29. Probability density versus mole fraction of CH ₄ , O ₂ , and N ₂ for array sizes of 1, 2, 3, and 4 for predicting 0.1 CH ₄ , 0.2 O ₂ , and 0.7 N ₂ . Results are from the best arrays all on the left, and subsequent arrays are evenly spaced based on decreasing performance (100 th , 66 th , 33 rd , 0 th percentile, ranking from left to right).	59
Figure 30. Plot of KLD (in bits ³ , averaged over 10 GA runs) versus population size, for selecting the best 4 MOF arrays. Dashed line is the KLD of the best possible 4 MOF array.	60
Figure 31. Plot of average KLD (in bits ³) with the range of possible values shaded versus the generation for selecting 4 MOF arrays with a population size of 50 and 100 arrays. The upper bound peaks at the KLD for the best possible 4 MOF array.	61
Figure 32. Plot of normalized values for KLD (in bits ³) and deviation in KLD versus the population size, for selecting 4 MOF arrays.	62
Figure 33. Plot of KLD values normalized, relative to the highest KLD for that array size, versus mutation rate in percent for 4, 10, and 25 MOF arrays.....	63
Figure 34. Plot of KLD values normalized, relative to the highest KLD for that array size, versus mutation rate, 0 to 1.4 percent, for 4, 10, and 25 MOF arrays. (Please note the difference in the y axes values from Figure 10.).....	64
Figure 35. Plot of <i>best</i> KLD value (in bits ³) versus array size, using a population size of 100 and a mutation rate of 0.1%.....	65
Figure 36. Plots of probability versus mole fraction of CH ₄ , O ₂ , and N ₂ for the best predicted arrays of a) 5, b) 10, c) 15, d) 20, e) 25, f) 30, g) 35, h) 40, i) 45, and j) 50 MOFs.	66
Figure 37. Schematic of experimental setup.....	70
Figure 38. Image of the MOF ZIF-8, chemical structure Zn(MeIM) ₂	72
Figure 39. Plot of frequency shift versus time for real time measurements of five gas mixtures of N ₂ , CO ₂ , and CH ₄	73
Figure 40. Negative frequency shift vs gas mixture for various concentrations of N ₂ , CO ₂ , and CH ₄	74
Figure 41. Truncated normal distribution for the Experiment 1 gas mixture in IRMOF-1	103
Figure 42. Truncated normal distribution for the Experiment 1 gas mixture in ZIF-8.....	104

Figure 43. Probability vs mole fraction for methane in each simulated mixture in IRMOF-1 ..	105
Figure 44. Probability vs mole fraction for methane in each simulated mixture in ZIF-8	105
Figure 45. Probability vs Mole Fraction for CH ₄ , O ₂ , and N ₂ predicted by an array of IRMOF-1 and ZIF-8.	107
Figure 46. Normal probability distributions for example MOFs 1 and 2	112
Figure 47. Probability versus concentration of methane for both example MOFs. Each point corresponds to a simulated gas mixture, for which probabilities were read from the normal distributions.....	113
Figure 48. Probability vs methane concentration for the array of MOF-1 and MOF-2.....	115
Figure 49. Probability versus concentration for CH ₄ , O ₂ , and N ₂ , as predicted by a two MOF array.	116
Figure 50. Probability versus concentration, including reference probability set Q (discrete version of Figure S3).	116
Figure 51. Plot of time in minutes versus array size for a population of 100 and mutation rate of 0.1.....	118
Figure 52. GA run time vs total MOFs to select from	119
Figure 53. Probability vs mole fraction for one concentration having essentially 100% probability.	120

Preface

There are many people that I would like to acknowledge, who contributed in innumerable ways to making the completion of this work possible. First, I would like to thank my PhD advisor, Dr. Chris Wilmer. He gave me an opportunity when I had no computational experience, and he has made me a better researcher and a more confident engineer. Next, I would like to thank Dr. Susan Fullerton, for listening when I needed to talk, and who has been an outstanding role model for me and many others in our department. I would also like to thank Dr. Beckman for offering a new, experimentalist's perspective on this work.

Next, I want to thank our experimental collaborators Dr. Paul Ohodnicki and Dr. Jagannath Devkota for validating our ambitious ideas and providing practical insights into sensor device construction. I also want to thank Paul for his encouragement, leading to my appointment at the National Energy Technology Laboratory.

This would not have been possible without the support of my fellow Wilmer Lab members; thank you to Alec, Kutay, Paul, Brian, Hasan, and Aigerim. I am grateful for the positive and warm environment you all created and for your willingness to always help. I would like to thank all my dear friends, near and far, especially Kate, Laura, and Ali for making Pittsburgh home, as well as Megan – the best person I know. Thank you to my family and my parents for always encouraging me and never holding me back. Most of all, I would like to thank my sister and best friend Breanne, for putting up with me.

This work began as an ambitious mission to develop a “universal” electronic nose, a gas sensor that could detect anything from methane leaks to lung cancer. However, upon more research into the field, Chris and I discovered that optimizing a general approach to developing improved

sensing technologies could have a significant societal impact in many distinct areas. While methane leak detection remains important, this work ultimately aims to optimize the design of a breath sensor to detect lung cancer and other diseases.

As the pioneer for this project in the Wilmer Lab, I developed methods that we believe are important for discovering the best possible electronic nose technologies. Using established molecular simulation techniques, we created two distinct metrics by which to compare MOF array configurations, followed by an algorithm to efficiently screen many different materials. In collaboration with experimentalists, we filed a patent for the device configuration, an array of surface acoustic wave devices- each with a different MOF material, in conjunction with our analysis methodology.

Additionally, we have presented our research with success in local business competitions. First, I placed second in the Start Up Blitz competition at Pitt, where I pitched a breath sensor for lung cancer detection. Then, Chris and I participated in the Pitt Innovation Challenge, a three round competition that includes a video submission, business and technical plan, and finally a formal business pitch. We were awarded a prize of \$100,000 for the continuation of this research.

This work includes two peer-reviewed articles, published in *The Journal of Physical Chemistry C* and *Sensors and Actuators B: Chemical*, and two more that have been submitted to *Sensors and Actuators B: Chemical* and *ACS Sensors*. Parts of the introduction will also be included in a review article that is in preparation, on the topic of computational developments of MOF-based sensors, to be submitted to a special issue of the journal *Sensors*.

1.0 A Brief History of Gas Sensing Technologies: Motivation for New Approaches and Improved Sensing Materials to Optimize Electronic Nose Design

1.1 An Introduction to Gas Sensors and Their Applications

Gas sensing technology has been actively developed for over a century, with notable successes in certain areas, such as residential hazardous gas monitoring (e.g. smoke detectors, carbon monoxide monitors) and quality control in manufacturing; however, *human olfaction* has been used in evaluating food quality (i.e. whether something was spoiled or rotten), identifying poisonous substances, and even for diagnosing diseases since ancient times.[1,2] In ancient Greece, physicians looked for the sweet smell of acetone in diabetics, a musty, fishy odor of advanced liver disease, a urine-like smell signaling failing kidneys, and a putrid scent from a lung abscess.[3] It was not until the 20th century that gas sensing became a field of major interest, leading to sophisticated sensing technologies like gas chromatography, metal-oxide-based sensors, optical sensors and more.[4]

There are many successful commercial gas sensors used in industrial food and product quality evaluation [2,5,6], processing controls,[4] indoor air quality assessment [5,7,8], and fire alarms. [5] Such sensing materials include metal-oxide semiconductors (MOS) and conducting polymers, devices like surface acoustic wave sensors (SAW), colorimetric and fluorescence (optical) sensors, in addition to classical analytical instruments such as gas chromatographs, electron capture detectors, flame ionization detectors, flame photometry detectors, infrared spectrometers, nuclear magnetic resonance spectrometers, mass spectrometers, and more.[4]

Despite an abundance of existing tools, there are still many applications for which current gas sensing technologies are inadequate. [1,2,5,9–13] For example, detection of diseases via breath or body odor,[5] low concentration methane sensing (to detect natural gas leaks) [14,15], and detection of toxic industrial chemicals in emergency response situations [16], where sensors need to be both accurate and portable [1,5].

Ideally, sensors would operate at room temperature, have minimal calibration requirements, longevity over many runs, short response times, and high stability. Additionally, specific applications like disease diagnosis require sensors to be portable, low cost, and have high selectivity. The most common commercial sensors are metal-oxide based, such as SnO_2 , due to their high sensitivity and reliability for many industrial applications. However, they must be operated at high temperatures from 300°C to 500°C, and such a high temperature requirement leads to large power consumption, as well as a lack of portability.[4]

Additionally, conductive polymer sensors have proven useful for applications such as fire detection, water pollutant sensing, and food product quality.[17] Their signal transduction occurs through changes in electrical resistance due to gas adsorption. These polymers exhibit high sensitivity, short response times, easy synthesis, and ideal mechanical properties; however, they are highly susceptible to humidity and their sensitivities are typically an order of magnitude lower than those for metal oxides.[2,4,17–19] Electrochemical gas sensors are able to operate at room temperature and have low power consumption, but they remain bulky. Moreover, their sensing mechanism is based on electrochemical oxidation or reduction of volatile compounds at a catalytic electrode surface; thus, they are not sensitive to a wide range of compounds, including aromatic hydrocarbons.[4]

Another class of sensors are optical-based sensors, including colorimetric and fluorescent sensing methodologies. First, colorimetric sensors employ films of dyes that respond to a variety of chemicals with changes in color, typically employed in arrays (discussed more in the next section).[\[4,20\]](#) These sensors are attractive due to their small size, low power consumption, and ability to target specific analyte types, but challenges arise of reproducibility, stability, and ability to distinguish specific molecules (i.e. they are better suited for analyte classification).[\[20,21\]](#) Similarly, fluorescent sensors detect light emissions from the gases of interest, which occurs at a lower wavelength than for colorimetric sensors and therefore they are more sensitive.[\[4,22,23\]](#)

Acoustic wave devices have been in commercial use since the early 1900s, but it wasn't until 1979 that their use for chemical vapor detection was first reported.[\[4,24\]](#) They are prevalent due to their small size (order of millimeters), their low manufacturing costs, and their sensitivity to a wide range of gases. These sensors can operate by either of two mechanisms: 1) a bulk acoustic wave (BAW) (as in a quartz crystal microbalance (QCM)) or 2) a surface acoustic wave (SAW), as shown in Figure 1.[\[4\]](#) The device platform consists of a layer of piezoelectric material, like quartz, where for BAW devices the quartz is the majority of the platform and for SAW devices the platform has the piezoelectric only as a thin surface layer.[\[25\]](#) Both devices are classified as microelectromechanical systems (MEMS) devices, meaning they convert a physical property change into an electrical signal. For example, upon exposure to a gas mixture a SAW device, as gases adsorb to the sensing layer the surface responds with a shift in frequency of the oscillating surface wave. The gases depositing onto the sensor surface cause a change in mass on the device and therefore a shift in the wave frequency. Such sensors are highly sensitive but depend on the sensing material used.[\[24–30\]](#)

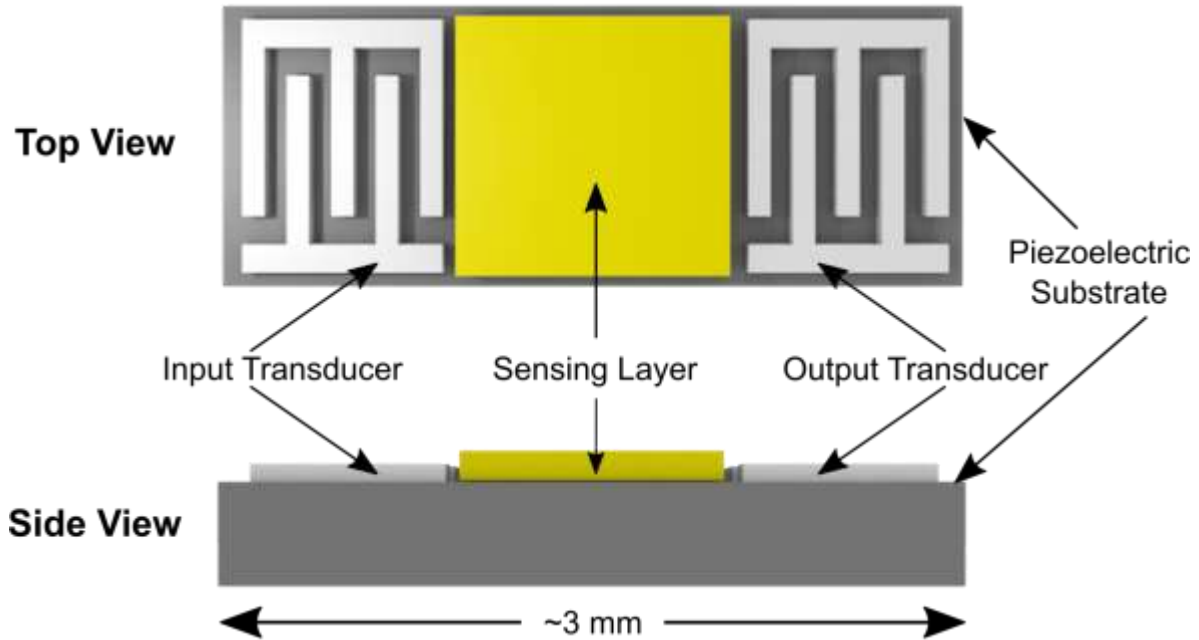


Figure 1. SAW device

Success has been reported for the use of ZnO nanorod SAW sensors for hydrogen detection, but the operating temperatures remain above 200°C; however, integration with metals like platinum have shown good results for detecting 1% hydrogen at room temperature.[28] Another example is ammonia sensing, for which the first SAW sensor was reported in 1987 using a platinum coating.[31] Semiconductor materials used with SAW devices include WO₃, showing good sensitivity towards NO₂, H₂S, O₃, and H₂. [28,32,33]

Additionally, we can report on current studies for SAW sensing as applied to a variety of gases. Methane sensing has been reported for lower limits below 500 ppm via SnO₂, but operation conditions are above 300°C; [34] cryptophane-A was able to detect below 0.05% at room temperature.[35] Some sensors have detected ppm levels of CO₂, consisting of Teflon-AF, spin-coated polymers, and functionalized single-wall carbon nanotubes.[25]

Commonly, polymers have been used on such QCM and SAW devices, as they are well known, have shown good sensitivity towards gases, and have a wide structural variety. Although they are adequate for their existing uses, most polymers cannot achieve the sensitivity and selectivity necessary to tackle medical and environmental challenges. For example, polypyrrole nanofibers were used as a sensing layer, detecting 1% hydrogen gas with a frequency shift of 20 kHz; however, the sensing mechanism for this is unclear and it is challenging to tune these materials for increased sensitivity.[36] A study with a bilayer structure, employing copper phthalocyanine with platinum layers, was sensitive to hydrogen in the range of 0.5 to 3%, promising for low concentration hydrogen detection.[37]

Despite the existence of polymer-based SAW sensors that exhibit refined sensing capabilities, this technology is still not able to resolve complex gas mixtures and the trial-and-error sensor development process is time consuming and costly. Exhaustive work has been done to maximize the sensing performance of polymers to target specific gases; thus, the optimal solution for advanced applications may rely on other sensing materials.

Recent studies have demonstrated novel gas sensing techniques that have the potential to provide early detection of ovarian, lung, and colorectal cancer through the analysis of a patient's breath, but reliability of those sensors, imperative for medical diagnoses, has room for improvement [1,38–42]. Hazardous chemical sensors [5,7,22] and environmental, as well as indoor air detectors [5], exhibit a multitude of problems that inhibit their use, such as lack of reproducibility and portability [2,16], low sensitivity and selectivity [5], long response times [5], and expensive device manufacturing and operation.[9]

1.2 Electronic Noses

Gas sensor *arrays*, also called electronic noses, are analogous to mammalian olfactory systems in that they employ multiple sensing elements to analyze complex odors. Human noses have only about one million receptors, compared to the 100 million present in dogs, allowing the animals to distinguish smells more effectively; moreover, dogs have been found to be able to distinguish patients with lung cancer.[1,4,42] Much work has been done to mimic the sensing capabilities of dogs noses, but no technologies have been able to achieve their sensitivity and selectivity for a wide range of applications.[42] In 1982, the first gas multi-sensor array was invented, and hundreds have been developed since.[43] In 1988, Gardner and Bartlett coined the term “electronic nose”, defining it as “an instrument which comprises an array of electronic chemical sensors with partial specificity and appropriate pattern recognition system, capable of recognizing simple or complex odors”. [4] Prior to this, though, others had published studies on olfaction devices, measuring aromas using conductivity and contact potential.

Electronic noses are already used in industries such as food, drug, and cosmetics for quality control and product monitoring.[4] As discussed in the previous section, there are a wide variety of sensing technologies, many of which can be used for array-based sensing, including metal-oxide, semiconductive polymers, conductive electroactive polymers, optical, surface acoustic wave, and electrochemical gas sensors.[4] However, the same issues exist for these sensing mechanisms as described in the previous section.

There are several electronic noses that have been in industrial use, which continues to grow as new applications are found.[4,44] Specifically, the Applied Sensor Company has successfully deployed their Air Quality Module electronic nose over the past ten years, which detects air quality including VOCs and carbon dioxide indoors.[4] Additionally, the Alpha-MOS Fox electronic nose

employs metal oxide gas sensors in combination with external carrier gas bottles in a flow-injection system.[4] The Cyranose 320 is portable and made up of polymer sensors blended with carbon black composite, used to identify eye infection causing bacteria with the combined use of three nonlinear analysis methods.[45] A handful of other commercial sensing technologies exist, but have been developed to target application specific gases and cannot be translated to more complex systems.

There are many potential applications whose technical requirements are still out of reach for current electronic noses. Detecting hazardous gases as a first responder, for example, is a challenge due to the large number of both target gases and possible interferents.[46] Similarly, detection of various diseases via analysis of the breath is beyond the capabilities of existing commercial electronic noses, although impressive progress has been recently reported.[1,38,40,47]

Another challenging problem is the detection of low concentrations of methane in air to detect natural gas leaks. Methane leakage is a significant contributor to global climate change and reported measurements may be inaccurate. [14,48] Current commercial technology cannot accurately read ppm or ppb methane levels using portable, low-power sensors.[49–51] Efforts to create better natural gas sensors have included metal-oxide catalysts, which are easily poisoned by halogens and do not survive extreme environmental conditions;[15,52] sensors with higher operating temperatures that have a high risk for explosions;[48] and traditional energy-consuming electrochemical sensors, where only minimal success has been achieved in low-cost portable sensor development.[53]

1.3 New Classes of Sensing Materials

Polymers and metal oxides have been, and continue to be, the most commonly used sensing materials in gas sensor devices.[2,5] Recently, other materials have been investigated for their use in gas sensing, such as carbon nanotubes (CNTs), [1,5,54,55] metal-organic frameworks (MOFs), [56,57] and zeolites, which are described here in further detail.[1,5]

1.3.1 Carbon Nanotubes and Zeolites

CNTs can be defined in two distinct types: multi-walled carbon nanotubes (MWCNTs) and single-walled carbon nanotubes (SWCNTs).[28] A SWCNT is a graphene cylinder with diameters ranging from 0.4 to 2 nm, where MWCNTs consist of nested cylinders of the same type.[28] CNTs exhibit high sensitivity (three to four times greater) relative to existing organic layer-coated sensors[29] and their electronic properties allow for compatibility with electrical signals associated with sensing technology. To increase the sensitivity towards NO₂, defects have been introduced into CNTs with success, and additional work has been done to dope and modify the structures to target specific compounds.[58]

However, CNT synthesis does not produce monodispersed materials with controlled physical and chemical properties, and they can only detect molecules able to distinctly accept or donate electrons. [5,55] Additionally, they have high recovery times, lower sensitivity than desired, and typically better signals are achieved at temperatures above 100°C.[28] In some cases, response times may take up to minutes, where seconds are desired.[58] While CNTs are a promising sensing material, much work must be done to tune the structures based on sensing application, especially for multicomponent mixture evaluation.

Another class of promising sensing materials is zeolites, due to their ordered 3D porous structures that allow for selective analyte adsorption.[59] Typically, modifications are made to the zeolites to create effective sensors, including zeolite-polymer composites, dispersion of zeolites into a conductive composite matrix, as well as chemical bonding of the zeolite on a substrate. Additional zeolite variants have been studied to improve sensor sensitivity and stability.[60] Despite their adjustable pore sizes and catalytic activity, issues of slow diffusion in the pores may become an important factor, where a tradeoff between selectivity and response time arises; moreover, their optimal operating temperatures are above 300°C.[61] As is the case with CNTs, there is a lot of trial-and-error work that remains if zeolites are to be used in commercial electronic noses.

1.3.2 Metal-Organic Frameworks

Although both MOFs and zeolites are nanoporous, crystalline, and have very high internal surface areas[5,56,57,59–62] that make them attractive sensing materials, MOFs are especially promising due to their reproducible synthesis and their easily tunable pore structures, ideal for targeting adsorption behavior to specific gas mixtures.[56,57,62,63] MOFs consist of organic linkers with metal centers and self-assemble in solution to form crystals. Despite being a relatively new class of materials, there have been significant strides in the discoveries of new structures.

For these reasons, several studies have explored MOFs as sensing materials, implemented as singular MOF-sensor devices.[57,64–67] However, only recently have MOFs been studied as component materials for electronic nose devices.[56,68] For example, Ellern et al. deposited MOFs onto microcantilever arrays and demonstrated their ability to detect gases.[69] Campbell et al. created an array of chemiresistive sensors, using conductive MOFs, to distinguish various

types of volatile organic compounds (VOCs).^[70] Another example is ZIF-8, a well-studied MOF known for its stability and small pores, which was employed by Yao et al. as a thin film sensing layer on arrays of metal-oxide nanowires to increase sensitivity.^[71] A recent study by Campbell et al. considered chemiresistive MOF-based sensor arrays that could distinguish sixteen different compounds into five categories: alcohols (methanol, ethanol, propanol), ketones/ethers (acetone, methyl ethyl ketone, tetrahydrofuran, dioxane), aromatics (benzene, toluene, p-xylene), aliphatics (hexane, cyclohexane, heptane), and amines (butyl amine, isopropyl amine, diethyl amine).^[70]

As mentioned above, MOFs may be used with a variety of signal transduction mechanisms and device platforms, including MEMS devices like SAW and QCM devices. Besides the benefits already mentioned of MOF-based sensors, they are exceptionally promising because they have shown to have record high gas adsorption,^[72–74] there are thousands of possible MOF structures that have not been explored for sensing, and their behavior can be modeled *in silico*. Importantly, this allows for computational screening of the materials for their performance in gas sensing environments.

There have been several computational studies for MOF gas sensing, which is the focus of this work. Specifically, Zeitler et al. have found trends in MOF properties for mass-based O₂ sensing in air, which have shown good uptakes at low pressure for MOFs with small pore diameters.^[75] Additionally, MOFs IRMOF-1 and HKUST-1 were tested via molecular simulations and it was concluded that the former is best for methane storage and the latter for methane/CO₂ separation.^[76] The same type of simulations were performed by Zeitler et al. for thin film sensing, modeling SAW, QCM, and microcantilever sensors. The adsorption of methane gases on the MOF coating were simulated using generic force fields, first screening for Henry's constants and isosteric heats of adsorption. It was found that atmospheric conditions like pressure

play a role in MOF performance, pore size plays a more important role than surface area or free volume, and Cu based MOFs are able to detect methane without the interference of water.[68] Lastly, Greathouse et al. simulated HKUST-1 and IRMOF-1 for the detection of aromatic molecules, where smaller molecules were able to pack more densely and had higher volumetric loading, but smaller hydrocarbons did not interact as strongly.[77]

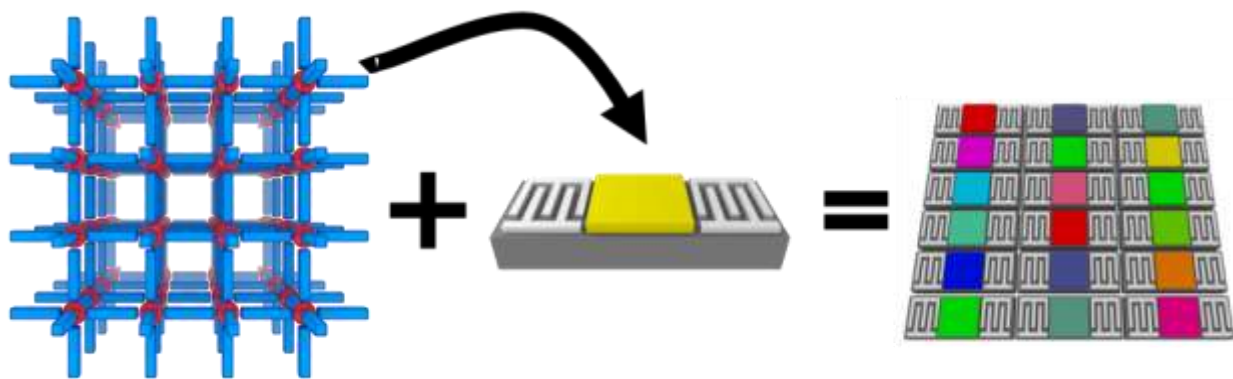


Figure 2. MOFs as sensing materials on SAW sensor array

All of these studies emphasize the importance of atomistic modeling to inform experimental sensor testing, so that efforts to develop optimal sensors are not wasted on poor performing candidate materials. While these studies are promising for the use of MOFs as sensing materials, they have been limited to a handful of MOF structures. Additionally, work has not been done to evaluate how they behave in array environments.

Sensor arrays have been proven to amplify performance, where diversity of sensing materials within the device provide more information in the output signal. Knowing this, we have computationally investigated the use of *different* MOF structures in an electronic nose configuration, which has not previously been done. In Figure 2, we have shown an example of a

MOF as the sensing material on a SAW device, where multiple sensors having different MOF sensing materials could be constructed to optimize electronic nose performance.

1.4 Computational Methods for the Analysis of Electronic Nose Signals

An essential component of an electronic nose is the interpretation of the output signal, which contains complex information from various array elements. In addition to testing different sensing materials in arrays, significant effort has been made in perfecting signal processing techniques to extract more information from fixed sensor array designs. These techniques include principal component analysis (PCA), linear discriminant analysis (LDA), partial least squares (PLS), artificial neural networks (ANN), and other cluster analysis techniques[[26](#)] to reduce dimensions in array data and compare classification abilities (given experimental results).[[1,4,7,10–12,21,26,78–81](#)] For example, common pattern recognition techniques like PCA take high dimensional input data and reduce it so that only the major contributing components are considered. Array data points are then clustered according to types of compounds (e.g. aromatics, hydrocarbons, etc...). While these methods work well in cases of classification and distinguishing single components, there is a need for more sophisticated methods to identify concentrations in complex gas mixtures.

Computational methods have been used to assess the effect of array size on electronic nose performance, with some interesting results. In 1998, a study by Ricco et al. concluded that, after testing SAW arrays of sizes three to six, that the optimal array size is 5-7 and any larger will cause a diminishing of device performance.[[26](#)] Their experiments employed various polymers and metal oxides, then processed using a pattern recognition algorithm.

Alkasab et al. have also employed information theory to analyze elemental contributions of sensor arrays and how they relate to the overall sensor's ability to provide information about the analytes of interest.^[82] These analyses are limited in that they use the number of correct analyte classifications as a scoring metric, they test only polymer and metal oxide sensing materials, and they require experimental sensor data to test performance. Over the last two decades, classes of sensing materials have expanded, but this limitation of array size has been accepted as a rule.

Although sensor arrays have been fabricated and studied and it is intuitive that increasing the number of sensors in an array *should* increase performance, *quantification* of that benefit by systematic variation of arrays, either in size or composition, has not been widely explored due to the significant cost of doing so experimentally. While prior reports suggest that larger arrays convey more information about the gas mixture, they do not directly compare different array designs and hence do not *quantify* the differences.^[11,13,26,54]

Our work aims to leverage the ability to computationally predict gas mixture adsorption in MOFs to rapidly and systematically screen different sensor arrays that vary in both the number of elements and in the type of MOFs used. It should be noted that even computational screening, though much faster than experimental testing, requires a tiered approach for finding optimal MOF arrays for sensing. Not only are there many thousands of MOFs to choose from, but there are many millions of ways in which those MOFs can be combined into an array. Therefore, what we propose is an inexpensive computational approach that can narrow the search space from millions of arrays to hundreds or thousands, followed by more expensive (and sophisticated) computations, and ultimately experimental testing.

2.0 Computational Design of Metal-Organic Framework Arrays for Gas Sensing: Influence of Array Size and Composition on Sensor Performance

Jenna A. Gustafson, Christopher E. Wilmer

J. Phys. Chem. C, vol. 121, pgs. 6033-6038, (2017).

Until now, there has not been an investigation into the use of different MOF structures as sensing materials in the form of arrays. Moreover, existing computational tools have not been used to systematically screen MOF array performance. This work aims to provide an approach for screening arrays of various sensing materials to optimize electronic nose performance, but also serves as a novel method for processing electronic nose signals to resolve complex gas mixture information.

2.1 Molecular Simulations of Gas Adsorption

Once a gas species adsorbs in a MOF, its presence can be detected in multiple ways, such as by measuring a change in mass or electrical conductivity, also known as the transduction mechanism [[56,57](#)]. In order to predict the ability of a sensing material, we first determine the overall transduction mechanism of that material when it is layered onto a device. Assuming the MOF would be layered on a MEMS device, we assume the transduction mechanism is related to the total mass change on the device, as exhibited in the Sauerbrey Equation (Equation 2-1). In

Equation 2-1, $\Delta m/A$ is the mass change per area, C_m is the mechanical coefficient for quartz, f_0 is the fundamental resonance frequency in reference state, and f is the output frequency shift.

$$\frac{\Delta m}{A} = \frac{C_m(f-f_0)}{2f^2} \quad (2-1)$$

In this study, we use a computationally inexpensive scoring metric, described in the following section, to quantify the potential of 31 different sensor arrays. We chose five different MOFs for this study: IRMOF-1,[\[83\]](#) HKUST-1,[\[84\]](#) NU-125,[\[85\]](#) UiO-66,[\[86\]](#) and ZIF-8.[\[87\]](#) These particular MOFs were chosen based on their diverse pore geometries and on the availability of experimental data. Below we describe in detail the method we developed to quantify sensor performance and compare the qualities of various MOF sensor arrays.

We considered gas mixtures with varying compositions of four different components: carbon dioxide, nitrogen, methane, and ethane. We chose these gases because of their relevance to natural gas-related sensing, among other applications, such as CO₂ measurement and air quality control, and because we had reliable force field parameters for each species. We used a set of gases and MOF structures with diverse chemical and structural properties to establish that our method is applicable for a range of gas compositions, and therefore will be potentially useful in high-throughput screening when developing any type of sensor.

Grand canonical Monte Carlo (GCMC) simulations were used to calculate adsorption of gas mixtures at 298 K and pressures of both 1 bar and 10 bar for each of the five MOFs studied. Ambient pressure sensing is relevant to many common applications (e.g., hazardous materials sensors [\[16\]](#), cancer detection [\[5\]](#), indoor air quality [\[5,7,8\]](#)) and higher pressure sensing is relevant

in the context of geological surveying [14] (e.g., detecting methane concentrations at various depths in oil wells). Universal Force Field (UFF) parameters were used for the MOF framework atoms, and the TraPPE force field was used for the gas species [88–91].

Regarding the accuracy of our predictions, we provide for context a comparison of our simulation data and prior experimental [73] adsorption measurements for all four gases in HKUST-1 (see Figure 3). Note that the comparison is between single-component simulations and measurements. The similarity of our simulation results to the experimental measurements in HKUST-1 gives confidence in the overall approach for rapidly screening MOF-based sensor arrays. However, this comparison also reveals that for sensing applications where simulated adsorption needs to be accurate to within 1% or less, significantly better force field parameters would be necessary.

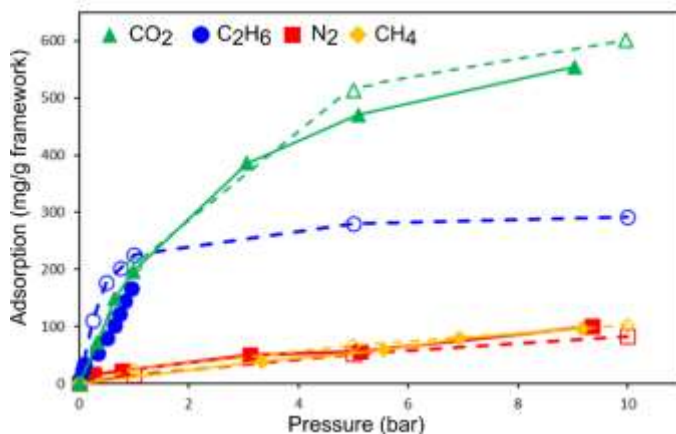


Figure 3. Adsorption (mg/g MOF framework) of CO₂ (green triangles),[92] C₂H₆ (blue circles),[76] N₂ (red squares),[92] and CH₄ (orange diamonds)[93] in HKUST-1 at 298K. Solid lines with filled symbols indicate experimental measurements taken from the literature and dashed lines with open symbols indicate our simulation predictions.

We considered 78 gas mixtures of the four gases, with N₂ ranging from mole fractions of 0.4-0.9, CH₄ from 0-0.575, CO₂ from 0-0.425, and C₂H₆ from 0.01-0.175. Please see Appendix B for full list of gas mixtures. We assume a signal transduction mechanism based on measuring the change in mass of a MOF due to adsorbing the gas mixture. After simulating adsorption for all gas mixtures in our five MOFs, we calculated the volumetric mass change (mg/cm³ of MOF) in each case, based on the assumption that the MOFs would be deposited as thin films of fixed thickness (rather than mass) on a sensor device (see Figure 4) [2,9].

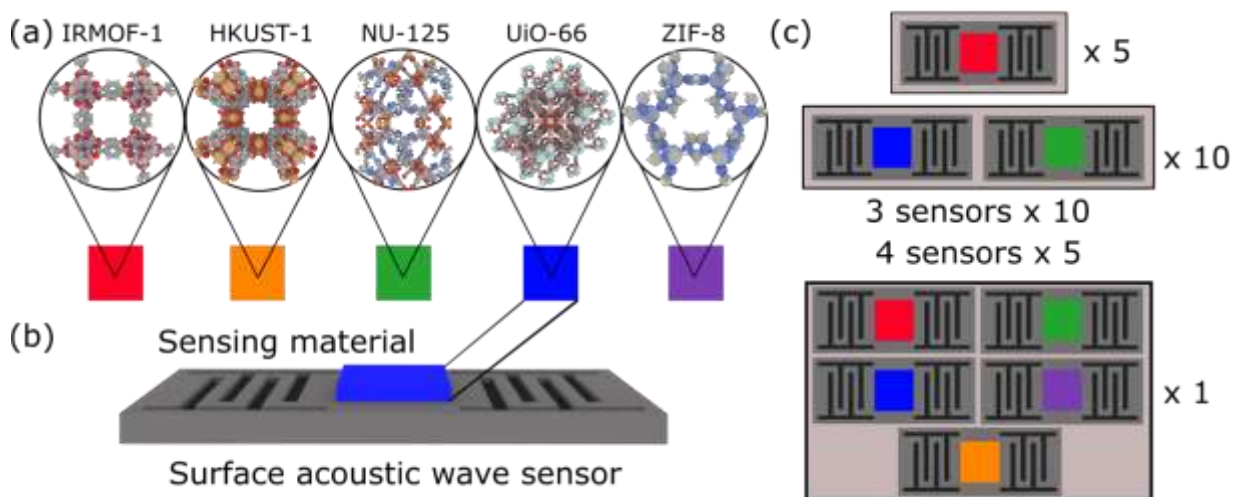


Figure 4. (a) The five MOFs we studied as sensing materials that could be (b) layered/deposited onto a surface acoustic wave (SAW) sensor. (c) Devices made from arrays of multiple SAW sensors on which different MOF structures are layered: 5 x 1 MOF arrays, 10 x 2 MOF arrays, 10 x 3 MOF arrays, 5 x 4 MOF arrays, 1 x 5 MOF array (31 arrays in total).

2.2 MOF Array Scoring Metric

Next, we considered various sensor array configurations based on all the combinations of the five MOFs. In order to quantify the potential of a particular MOF array for sensing applications, we have defined a simple performance score for each MOF-sensor array, called the *sensor array gas space* (SAGS) score, Φ . We hypothesize that the SAGS score will allow for the comparison of different sensor array configurations, leading to more insights on array design.

This score has the property that it is high for sensor arrays that have very distinct mass responses between gas mixtures that are similar in composition, and low for arrays that have similar mass responses when the gas compositions are very different. To calculate an array's SAGS score, we first calculate a *pairwise array* (PA) score, S_{ij} ,

$$S_{ij} = \frac{m_{ij}}{d_{ij}} \quad (2-2)$$

where d_{ij} is the Euclidean distance between two different gas compositions, i and j , each with N component gases, specified by their mole fraction, x_k , as shown in Equation 2-3,

$$d_{ij} = \sqrt{\sum_{k=1}^N (x_{k,i} - x_{k,j})^2} \quad (2-3)$$

and m_{ij} is the Euclidean distance between the mass changes in an M element MOF array adsorbing either gas mixture i or gas mixture j , as shown in Equation 2-4.

$$m_{ij} = \sqrt{\sum_{k=1}^M (m_{k,i} - m_{k,j})^2} \quad (2-4)$$

We expect the pairwise array score to indicate *how well* a MOF array can distinguish between a pair of gas mixtures. To calculate the SAGS score, we calculate the pairwise array score over all pairs of gas mixtures in a given space of gas mixtures, and then take the average (see Equation 2-5),

$$\phi_W = \frac{\sum S_{ij}}{W} \quad (2-5)$$

where W is the total number of combinations of pairs of gas mixtures used in the average (78 mixtures yield 3,003 pairs in this study).

A high ϕ_W means that, over the range of gas compositions considered, the array is good at distinguishing between very similar mixtures. Each combination of MOFs in the array has its own ϕ_W for a particular set of gas mixtures.

2.3 Ranking MOF Arrays

After calculating SAGS scores for all 31 possible MOF sensor arrays, we looked at both the effect of the size and composition of the array (see Figure 5). As the number of MOFs in the sensor array increases, the SAGS score increases. As expected, we found that larger arrays of sensors are better able to distinguish between gas mixtures of similar composition.

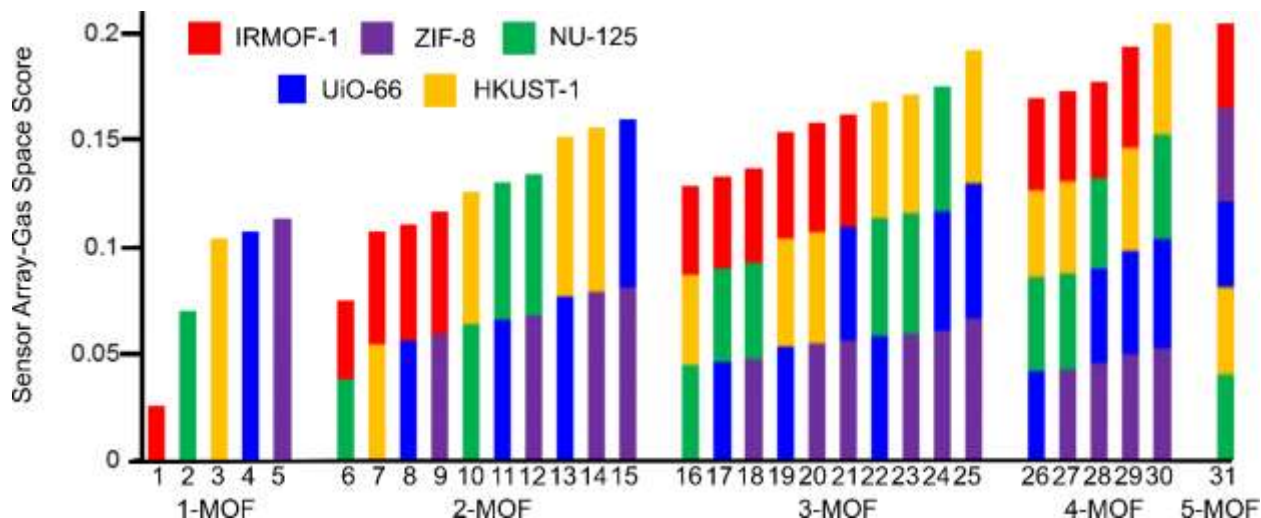


Figure 5. Sensor array gas space scores for MOF arrays consisting of different combinations of one to five MOFs in an array, using IRMOF-1, HKUST-1, NU-125, UiO-66, and ZIF-8 at 298K and 1 bar.

For example, at 1 bar, the five-MOF array, made up of HKUST-1, NU-125, UiO-66, IRMOF-1, and ZIF-8, resulted in a score of 0.205, while the best scoring three-MOF array, consisting of HKUST-1, UiO-66, and ZIF-8 had a score of 0.192. However, as more MOFs were added, each successive one had a smaller effect on the SAGS score. If the cost of a sensor-array increased linearly (or faster) with the number of elements, the diminishing growth of the SAGS score would imply that there is an optimal array size for a given application.

Further, while there is a large difference in the score between the best and worst MOF among the 1-MOF “arrays,” the gaps between the best and worst arrays of larger sizes are relatively smaller. Notably, this is because the worst 1-MOF sensors become significantly better when other MOFs are added to them. Although a single MOF can have a very low score (e.g., IRMOF-1: 0.025), the score of the worst pair was more than double (IRMOF-1 and NU-125: 0.075). Therefore, when designing new MOFs for gas sensing application, it may be easier to find *two* that work well together than to find *one* with high performance.

For the simulations of adsorption at 10 bar, we found that the SAGS scores were generally higher than at 1 bar, which is expected due to increased adsorption (and hence greater mass response) at larger pressures (see Figure 6). We found no correlation between the arrays that performed best at 1 bar vs. 10 bar. This simple observation highlights the fact that optimal *sensor arrays are unique to their operating conditions*; the optimal array at 1 bar is not necessarily the best one at higher pressures.

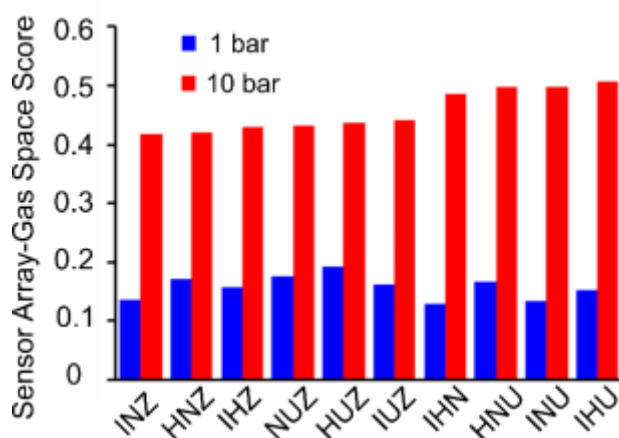


Figure 6. *Sensor array gas space scores at 298K, 1 bar and 10 bar, for arrays with three elements using combinations of IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), and ZIF-8 (Z). The data at 10 bar is ordered by increasing score along the x-axis, but the corresponding scores for sensors at 1 bar do not follow the same trend.*

We also analyzed the degree to which specific MOFs would improve the SAGS score if they were added to existing arrays of two or three MOFs (see Figure 7 and Figure 8). At 1 bar, Figure 7a shows that MOFs UiO-66 and ZIF-8 appeared to increase the SAGS scores the most when added to an existing two-MOF array.

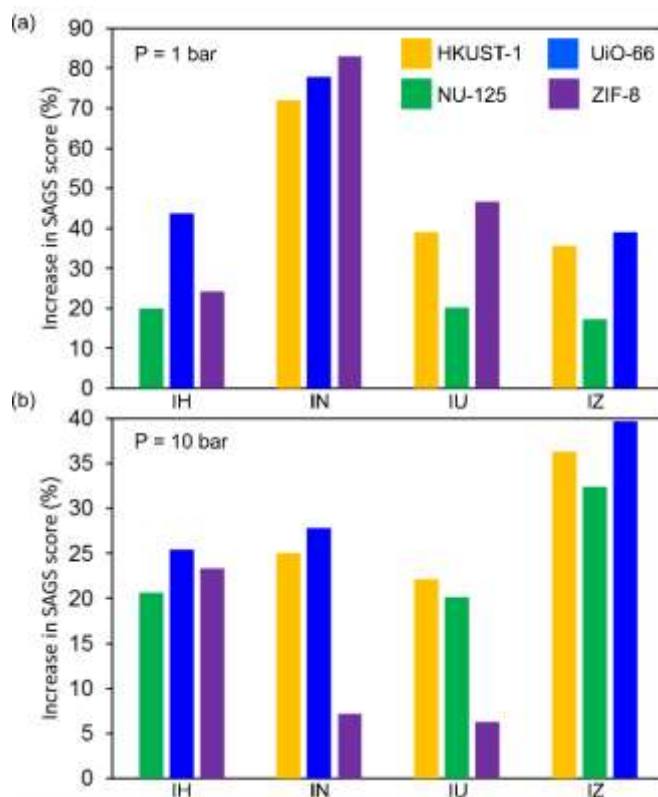


Figure 7. (a) Percent increase in SAGS score as one MOF (specified by color) is added to an existing two-MOF array, specified on the horizontal axis, at 1 bar: IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), ZIF-8 (Z). (b) Percent increase in SAGS score as one MOF is added to an existing two-MOF array at 10 bar.

We compare this result for the same configurations at 10 bar (Figure 7b), where instead UiO-66 and HKUST-1 increased the scores the most. Analogously, when beginning with a three-MOF array, at different pressures, different MOFs will have the greatest impact on the SAGS score: UiO-66 and ZIF-8 at 1 bar and UiO-66 and HKUST-1 at 10 bar, as shown in Figure 8a and Figure 8b, respectively.

Interestingly, in all cases at 1 bar, NU-125 was the least valuable MOF to add to an array based on its ability to increase the SAGS score. However, in all cases but one at 10 bar, ZIF-8 was the least valuable MOF to add to an array, with this singular exception being a two MOF array

consisting of IRMOF-1 and HKUST-1. Notably, in this exceptional case both MOFs have high void fractions, which typically implies poor adsorption at low pressures, such that this particular array might benefit complementarily from the relatively smaller void fraction of ZIF-8 and the correspondingly stronger van der Waals interactions in its pores. This further supports our observation that optimal MOF sensor arrays are those where each element has complementary adsorption behavior.

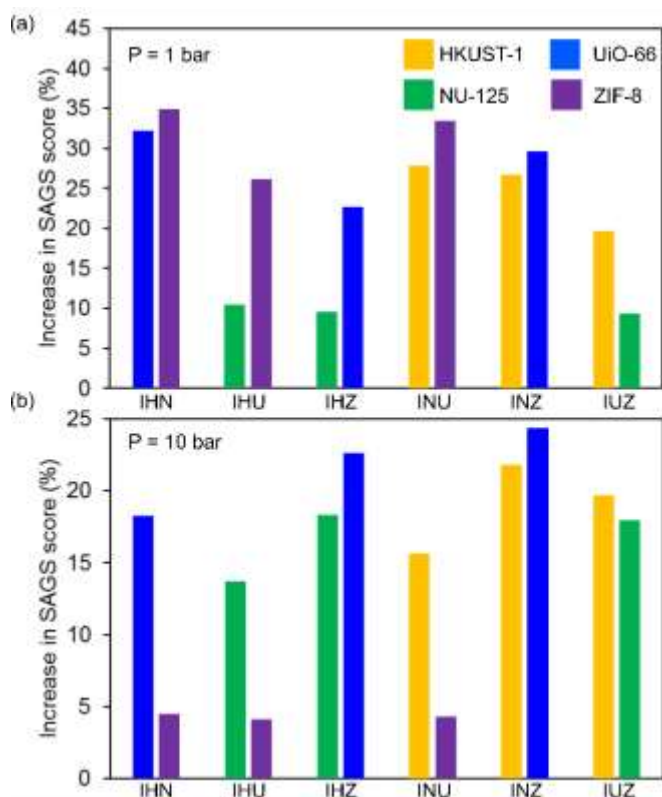


Figure 8. (a) Percent increase in SAGS score as one MOF (specified by color) is added to an existing three-MOF array, specified on the horizontal axis, at 1 bar: IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), ZIF-8 (Z). (b) Percent increase in SAGS score as one MOF is added to an existing three-MOF array at 10 bar

2.4 Effect of Structure-Property Relationships on Sensing

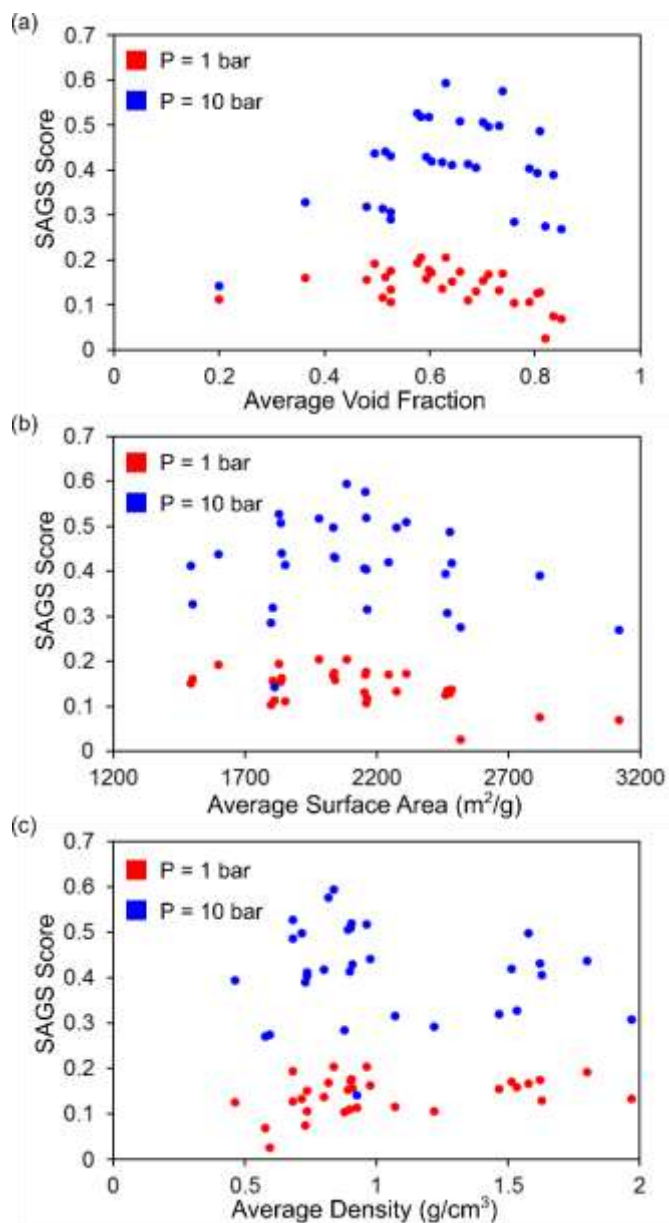


Figure 9. (a) SAGS score of each MOF array versus the void fraction averaged over all MOFs in the array, at 1 bar (red) and 10 bar (blue) (b) SAGS score versus the surface area averaged over all MOFs in the array, in m^2/g (c) SAGS score versus the density averaged over all MOFs in the array, in g/cm^3 . Please see Table 4 in the Supporting Information for a detailed list of MOF properties.

We also analyzed average physical properties of MOF arrays, namely void fraction, surface area, and density, to find simple structure-property relationships relevant to sensing (see Figure 9).

Complementarity aside, our findings show that for the gases tested, optimal average void fractions were at values of between 0.6 and 0.7 (see Figure 9a). The relationship between the SAGS score and the average void fraction is stronger than for the other two properties. We also observed, at 10 bar, that average surface areas of $\sim 2,000 \text{ m}^2/\text{g}$ led to the highest SAGS scores, but the peak was less pronounced than in the case of void fractions, and was not visible at all for the 1 bar data (see Figure 9b). We could not discern any direct relationship between SAGS scores and MOF densities (Figure 9c), which was expected. Though the number of data points in our study is too limited to draw strong conclusions about the relationship between sensing performance and physical properties, the relationships in Figure 9 indicate directions that can be explored in future work.

2.5 Conclusions

Systematically comparing different sensor arrays experimentally is highly nontrivial due to the significant cost of building multiple devices in a reproducible way. Here, using classical molecular simulations to model the adsorption of many four component gas mixtures in five different MOFs, at 1 bar and 10 bar, we have systematically evaluated 31 MOF sensor arrays based on a scoring metric that we defined called the SAGS score. This score quantitatively compares MOF arrays based on differences in total mass adsorption in each MOF over all of the gas mixtures considered.

As hypothesized, SAGS scores increased with sensor array size, but with diminishing returns, and the best arrays at 1 bar were not the same as those at 10 bar. Specifically, we determined that at 1 bar, UiO-66 and ZIF-8, and at 10 bar, UiO-66 and HKUST-1, had the most substantial effect on increasing the SAGS score when added to both two-MOF and three-MOF arrays. However, at 1 bar, NU-125 had the least effect on SAGS scores, while, on the other hand, at 10 bar ZIF-8 typically had the least effect on increasing SAGS scores.

Although in some ways unsurprising, the increased ability to resolve gas mixtures from combining multiple MOFs seems overlooked as a strategy in the context of the many studies that have focused on finding a single MOF that could be sufficiently sensitive and selective [56]. When faced with a MOF that is not sufficiently selective on its own, it may prove to be simpler to merely combine its signal output with that of another (equally mediocre) MOF than to improve the properties of either one.

It is important to note that the SAGS score is based on averaging behavior over a specified range of gas mixtures. An array may be excellent at distinguishing gas mixtures in a narrow range but still receive a low SAGS score when considered against a broader range of mixtures. Therefore, it is important to carefully choose the range of gas mixtures upon which the score is based to reflect what is expected to be relevant in a target application. For example, detection of natural gas leaks, one should choose gas mixtures of methane in air at sparing concentrations. Subsequently, after determining the score for many different MOF array combinations, one could select arrays with the highest SAGS scores and further analyze them using more rigorous methods.

Therefore, while the SAGS score can provide useful insights into array performance, we intend for it to be only one factor considered in the design of a sensor array. Nevertheless, when selecting an optimal sensor array from potentially millions of possible MOF combinations, the

SAGS score can serve as a useful high-throughput screening metric. When choosing the components of a sensor array from a smaller number of choices, we propose using a variety of other more sophisticated techniques, including principal component analysis, neural networks, and other statistical methods that are less well-suited for large-scale screening.

3.0 Optimizing Information Content in MOF Sensor Arrays for Analyzing Methane-Air Mixtures

Jenna A. Gustafson, Christopher E. Wilmer

Sens. Actuators B Chem., vol. 267, pgs. 483-493, (2018)

3.1 Background and Motivation

As previously mentioned, while there has been a focus on perfecting electronic nose analysis tools, there exists no efficient approach to select arrays of MOFs for a sensor application. Thus, a central problem in MOF-based electronic nose design has not yet been addressed: how does one choose an array of MOFs that would maximize device performance? Given that there are thousands of MOFs to choose from, and many millions of possible MOF-array combinations, it is unlikely that one would determine the most effective array by experimental trial-and-error. We propose a novel method to computationally screen MOF sensor arrays to determine which ones are best at resolving gas mixtures within a specified composition range.

Our approach differs from other commonly used methods to identify gases from electronic nose data, which are typically based on fingerprints, training data, and pattern recognition.[\[10,12,26,94\]](#) Electronic nose performance is often scored based on a sensor's percentage of correct classifications of compounds.[\[26,94\]](#) Common classification techniques include principle component analysis (PCA) or partial least squares-discriminant analysis (PLS-DA) with leave-one-out training, which reduce data dimensionality and group responses. Others

have used information theory measurements to statistically quantify the ability of individual elements to discriminate odors and compare to the overall array performance.[82] In contrast, we hypothesize that using a molecular model of gas mixture adsorption, instead of fingerprints, will efficiently extract detailed gas composition information from sensor signals.

As with the previous study,[95] we use atomistic simulations to predict gas mixture adsorption into each of the MOFs considered, which is assumed to cause a detectable mass change, measurable by a suitable device. Although our methodology is general with respect to signal transduction mechanisms, we nevertheless assume that a change in mass is measured via a MEMS device. For example, surface acoustic wave (SAW) devices are commonly used MEMS devices that measure small changes in mass and are compatible with MOF materials.[56] Then we reverse the process by *assuming* a change in mass is detected and “ask” each MOF sensor element to guess which gas mixture caused that mass change, resulting in a probabilistic distribution of possible gas mixtures.

By combining information from multiple MOFs in a candidate array (i.e., joining the probability distributions), we expect to accurately show the estimated gas composition from the eyes of each MOF array considered. Arrays that can more precisely resolve the gas mixtures—i.e., whose joint probability distributions have sharper peaks—result in higher Kullback-Liebler divergence (KLD) scores, which are rigorous measures of signal information content. Such a systematic analysis for ranking materials’ performances for electronic nose fabrication has not been developed before. An overview of the methods described in this section can be seen in the flow diagram, Figure 10.

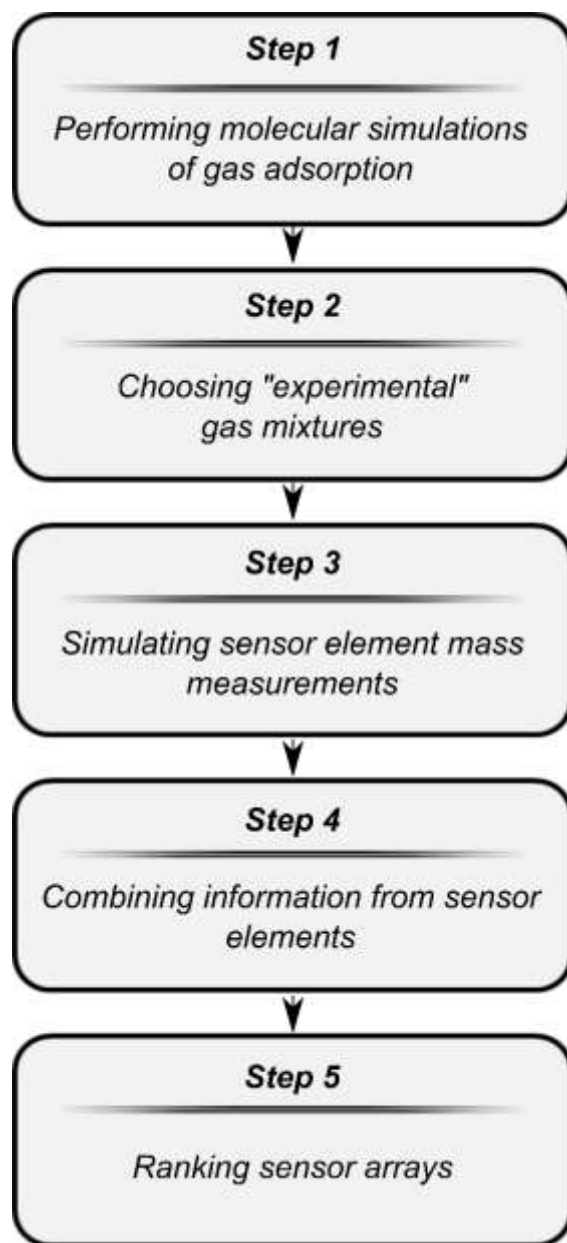


Figure 10. Flow diagram of analysis described in the Methodology section.

3.2 Performing Molecular Simulations of Gas Adsorption

Similar to our previous work,[\[95\]](#) grand canonical Monte Carlo (GCMC) simulations were performed using the RASPA software package[\[96\]](#) to obtain gas mixture adsorption data for a range of MOFs at 298K and 1 bar. The MOF structures used were IRMOF-1,[\[83\]](#) HKUST-1,[\[84\]](#) NU-125,[\[85\]](#) UiO-66,[\[86\]](#) ZIF-8,[\[87\]](#) MgMOF-74,[\[97\]](#) NU-100,[\[98\]](#) MOF-177,[\[99\]](#) and MOF-801.[\[100\]](#) We chose these particular MOFs because they are well-studied, have structural and chemical diversity, and are straightforward to synthesize in the lab. Unit cells of these MOFs are shown in Figure 11.

Simulation results of binary mixtures of N₂ and CH₄ are reported, followed by ternary mixtures of CH₄, N₂, and O₂, chosen as a simplified model for methane in air. Mixtures cover a range of 0-1 mole fraction for each component, in steps of 1%, resulting in a total of 5,151 mixtures for all ternary mixtures (of which the binary mixtures we refer to are a subset). For each gas mixture-MOF pair, the change in total mass (mg/cm³) due to adsorption is obtained from the simulation results.

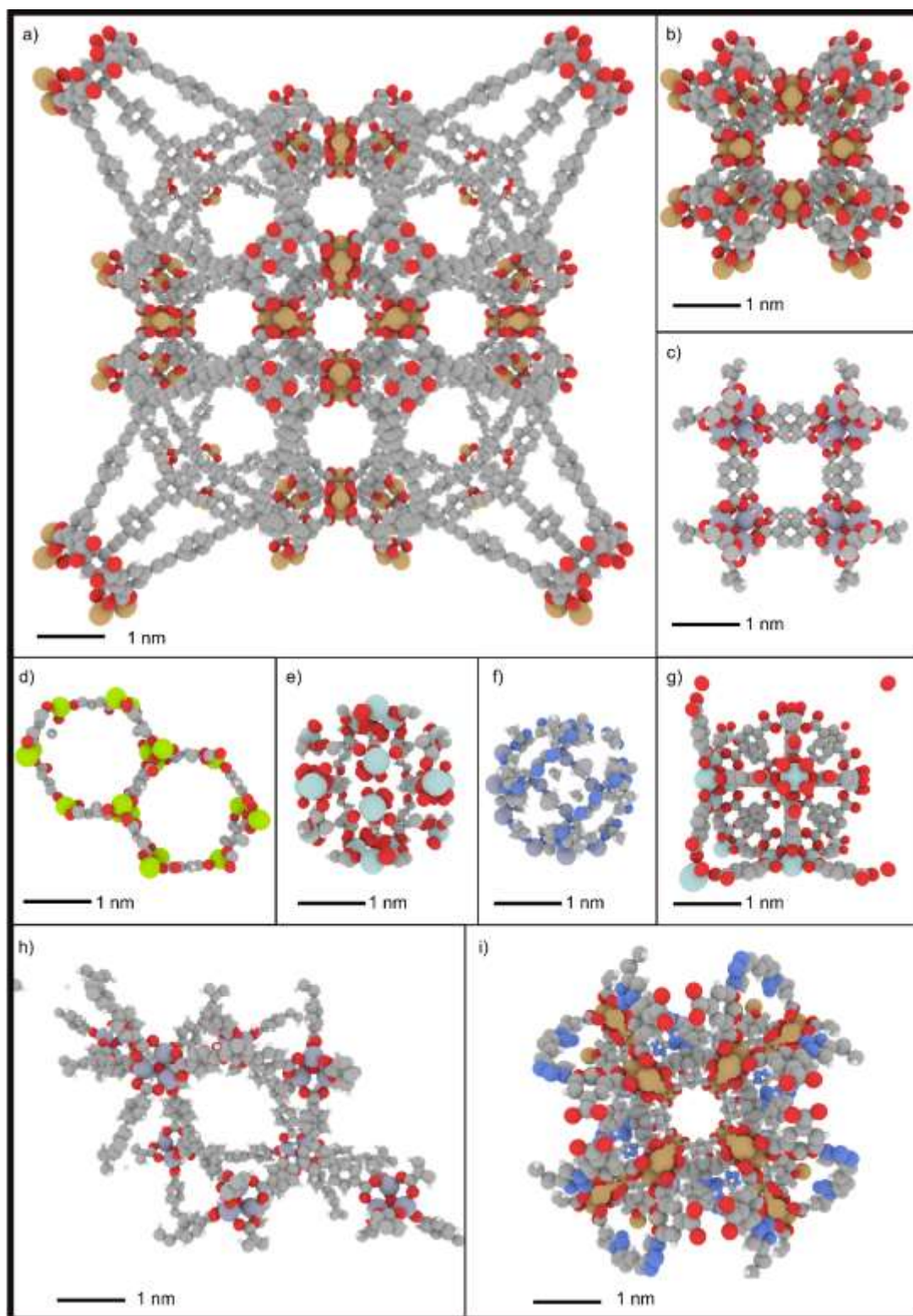


Figure 11. Unit cells of all MOFs used in this study: a) NU-100, b) HKUST-1, c) IRMOF-1, d) MgMOF-74, e) MOF-801, f) ZIF-8, g) UiO-66, h) MOF-177, and i) NU-125.

3.3 Choosing “Experimental” Gas Mixtures

To quantify sensor array performance, we performed simulated “experiments,” where arrays were exposed to unknown, or “experimental,” gas mixtures. Sensor arrays were then ranked on their ability to correctly estimate the composition of these unknown gas mixtures. In this step, we chose a few gas mixtures of interest from our simulated data set generated in Step 1 (Figure 10), and removed them from the data set, similar to leave-one-out regression methods. Specifically, we chose four binary mixtures of CH₄ and N₂ (Table 1), and nine ternary mixtures of CH₄, N₂, and O₂ (Table 2). We chose these compositions because we wanted to test sensor array performance over the whole range of concentrations for all gases. For each experimental mixture, we saved only the corresponding total mass adsorbed in each MOF, pretending the detailed compositions of the gas mixtures were unknown to us.

3.4 Simulating Sensor Element Mass Measurements

Once gas adsorbs on the MOF sensing material, a MEMS sensor element, such as a SAW device, can measure the corresponding change in mass. In this step, we simulate this mass measurement process by taking the total adsorbed mass calculated from our simulations and adding to it normally distributed noise to represent measurement error with a width equal to that of the device’s assumed precision, in this case 0.1 mg/cm³. Because adsorption always results in a positive mass change, we used a truncated normal distribution that always gives a zero probability for any negative values of mass.

Thus, we imagine a MOF sensor element is exposed to an unknown gas mixture, and then we compare its simulated mass measurement to all the other known mass results from our molecular simulations in Step 1. Intuitively, the gas mixture whose adsorbed mass is the closest to the simulated mass measurement is the most likely composition. However, many other gas mixtures may correspond to similar adsorbed masses that are within the assigned measurement error, resulting in a probabilistic distribution of possible gas compositions.

We obtain these probability distributions for every unknown gas mixture chosen in Step 2, for every MOF sensor element. Each probability distribution quantifies the likelihood, based on total mass change observed by an individual sensor element, that one of the previously simulated gas mixtures *is* the unknown gas mixture. The result is a set of 5,150 (one for each possible gas mixture) *discrete* probabilities for each MOF-unknown mixture combination for all nine MOFs. Please see Appendix D for a detailed example calculation.

3.5 Combining Information from Sensor Elements

We emulate a sensor array by joining the probability distributions of each element, calculated in Step 3. Here, a joint probability is calculated by multiplying n discrete probability distributions, for a total of n array elements, and then normalizing so that all of the points add up to one. Next, we test all possible sensor configurations using the nine MOFs, ranging from single element arrays (of which there are nine configurations) up to nine element arrays (of which there is only one configuration). For every MOF-array, the combined information result is still a list of 5,150 probability values, one for each of the possible $\text{CH}_4/\text{O}_2/\text{N}_2$ gas mixtures. We can then rank

each MOF array's performance by measuring the information content contained in the combined probability distribution.

3.6 Ranking Sensor Arrays

Thus, rather than plotting probability curves and visually comparing MOF arrays, we calculated the amount of information (in *bits*) provided by each array for detecting *a gas component* via the Kullback-Liebler divergence (KLD), as shown in Equation 3-1. The probability at each mole fraction is represented by P_i , and a reference probability of Q_i is a probability equivalent to $1/N$ (i.e., a uniform probability distribution), where N is the number of points (mole fractions) we have from 0 to 1 for each gas.

$$KLD = \sum_i^N P_i \log \frac{P_i}{Q_i} \quad (3-1)$$

This KLD value determines the *information content* of a probability distribution produced by an array, where a higher value is better. Thus, we rank arrays per their KLD values for the various “experimental” gas mixtures chosen in Step 2.

We demonstrate how this method can be applied to both binary (CH_4/N_2) and ternary ($\text{CH}_4/\text{N}_2/\text{O}_2$) gas mixtures using the same set of MOF sensor arrays. Predicting two-component gas mixtures is simpler, allowing for us to clearly validate our method. We expect this computational tool to facilitate efficient sensor array design, providing both a way of analyzing real sensor output measurements and optimizing materials selection.

GCMC simulations were performed on mixtures of CH₄, N₂, and O₂, for which a subset was particularly relevant to natural gas detection. The predicted total mass uptake for each MOF from the GCMC simulations is shown in Figure 12 for binary mixtures, while Figure 13 includes responses for three components. We see that mass adsorption (mg/cm³) for all MOFs increases linearly with increasing concentration of methane, with some MOFs exhibiting higher slopes, indicating a stronger methane selectivity, particularly for ZIF-8, MgMOF-74, and HKUST-1.

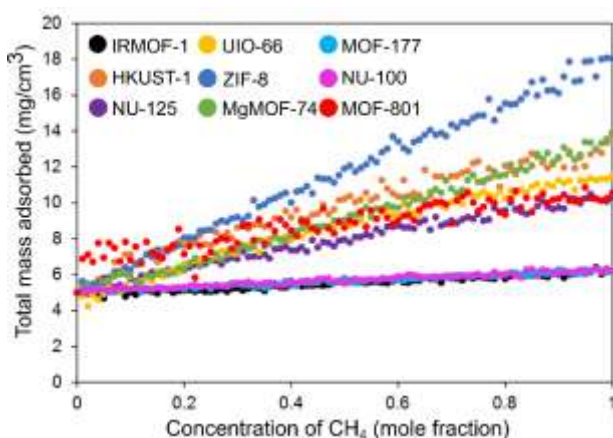


Figure 12. Total mass adsorbed (normalized by subtracting a value from each data set so all start at 5 mg/cm³ for a CH₄ concentration of 0) for nine MOFs vs concentration of CH₄ in various CH₄/N₂ mixtures.

The adsorption trends in Figure 13 provide insights into the individual gas behaviors for each MOF. All MOFs we considered exhibited a linear increase in mass with CH₄ concentration, although some gradients were steeper, as for NU-125, UiO-66, ZIF-8, and MgMOF-74. Additionally, we saw higher N₂ concentrations result in a slight linear decrease in mass for MOFs IRMOF-1, HKUST-1, NU-125, UiO-66, and NU-100.

Finally, adsorption decreased slightly with increasing O₂ concentration for MOFs UiO-66, ZIF-8, and MgMOF-74. There was essentially no impact of N₂ on the total mass in MOF-801, as

well as for O₂ in IRMOF-1, NU-100, and MOF-177. Per these adsorption trends, we hypothesized that MOF arrays would more accurately resolve concentrations of CH₄ and N₂ in the mixture than O₂, and we prove this to be true in the following analysis.

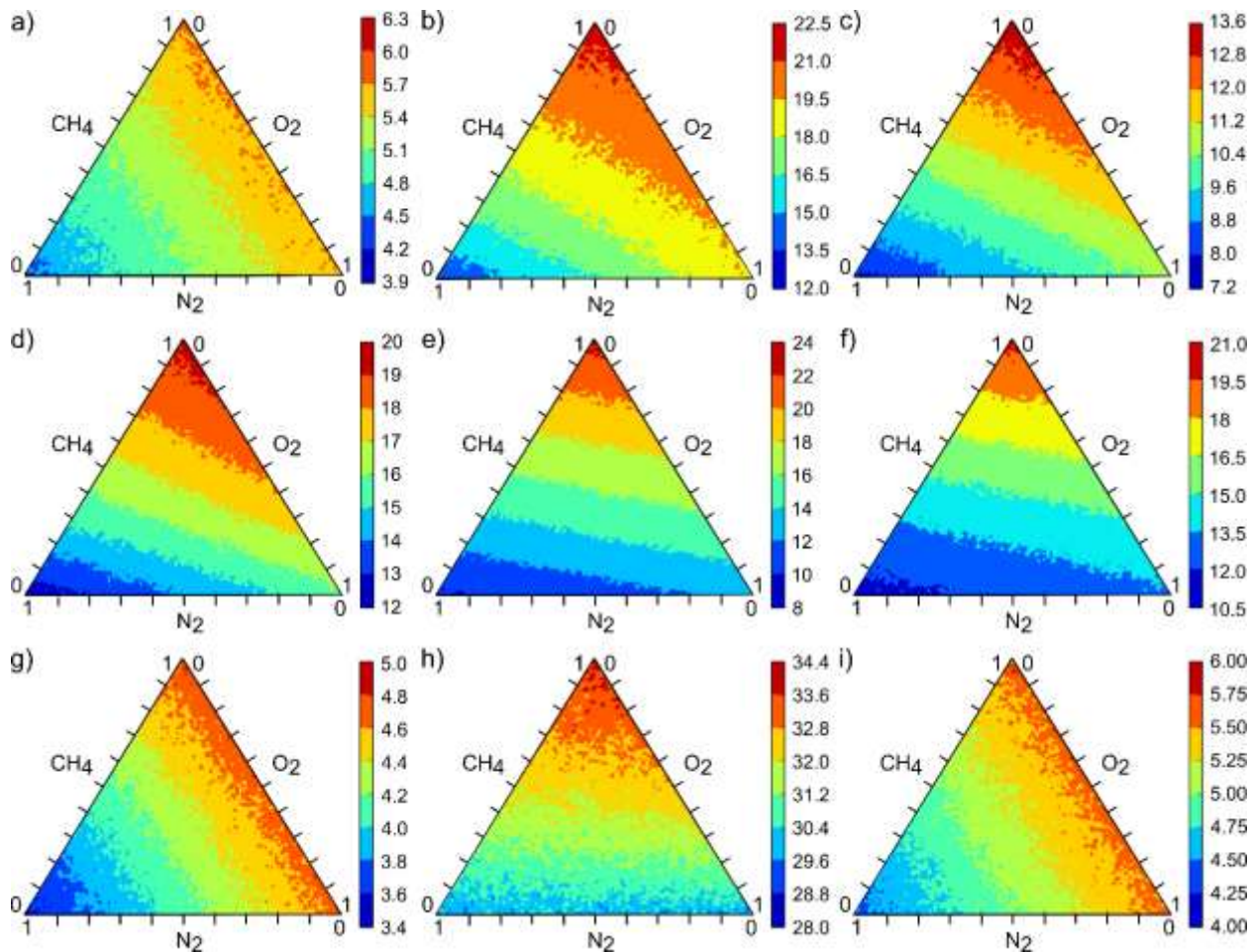


Figure 13. Total mass adsorption (mg/cm³) vs. mole fraction of various ternary gas mixtures in nine MOFs: a) IRMOF-1, b) HKUST-1, c) NU-125, d) UiO-66, e) ZIF-8, f) MgMOF-74, g) NU-100, h) MOF-801, i) MOF-177.

Note: The color bar scales vary between plots. Color represents the total mass of gas adsorbed (mg/cm³), and each triangle edge is the component mole fraction.

Table 1. List of binary experiments tested with mole fractions of each gas.

Experiment #	Component Mole Fraction	
	CH ₄	N ₂
1	0.1	0.9
2	0.25	0.75
3	0.5	0.5
4	0.75	0.25

Table 2. List of ternary experiments tested with mole fractions of each component.

Experiment #	Component Mole Fraction		
	CH ₄	O ₂	N ₂
1	0.1	0.15	0.75
2	0.2	0.2	0.6
3	0.3	0.3	0.4
4	0.4	0.2	0.4
5	0.5	0.2	0.3
6	0.6	0.3	0.1
7	0.7	0.1	0.2
8	0.8	0.1	0.1
9	0.9	0.05	0.05

As described in Section 3.4-3.5, we calculated the probability outcomes for each experiment to test whether an array could predict the concentration values. We analyzed all combinations of the nine MOFs listed above, by using four binary and nine ternary unknown gas mixtures, as listed in Table 1 and Table 2, respectively. The result of each experiment was a ranking of MOF arrays per their KLD values. We report the KLD values for *best* and *worst* performing arrays of each size, averaged over the four binary mixture experiments, as seen in Figure 14, below.

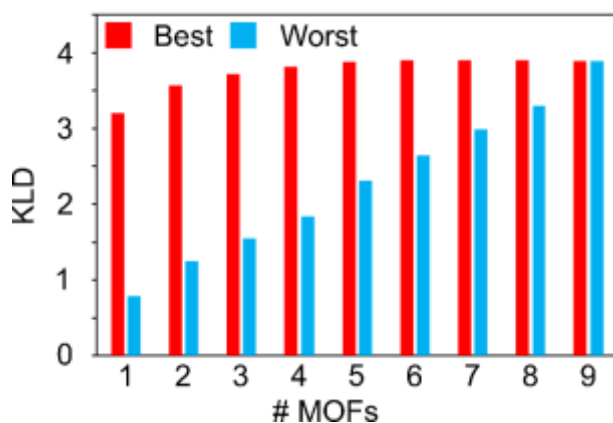


Figure 14. Average KLD vs number of MOFs in an array for the best (red) and worst (blue) arrays of each size over all four binary mixture experiments.

Focusing on the four-MOF case in Figure 14, we see that the *best* array performed much better than the *worst*, at KLDs of 3.81 and 1.84, respectively. This disparity in KLD values highlights the benefits of computational array design for gas sensing. From nine possible MOF materials, there are 126 possible configurations of four-MOF arrays. Thus, it is unlikely that the best array would be selected through a trial-and-error process, where synthesis and testing is time consuming; computational screening can significantly expedite the selection process.

Ultimately, the goal is to be able to choose the array that best fits the needs of a specified application, taking into consideration the time and resources needed to construct larger arrays. A four-MOF array may perform just as well as an eight-MOF array, with the right materials.

For all ternary gas mixtures, we looked at probability distributions of the concentrations of each *component* in the mixture; therefore, we calculated KLD values on a per component basis. Figure 15a shows KLD values averaged over all nine experiments, where at each array size, the three bars represent the KLD value for each gas component (CH₄, N₂, O₂).

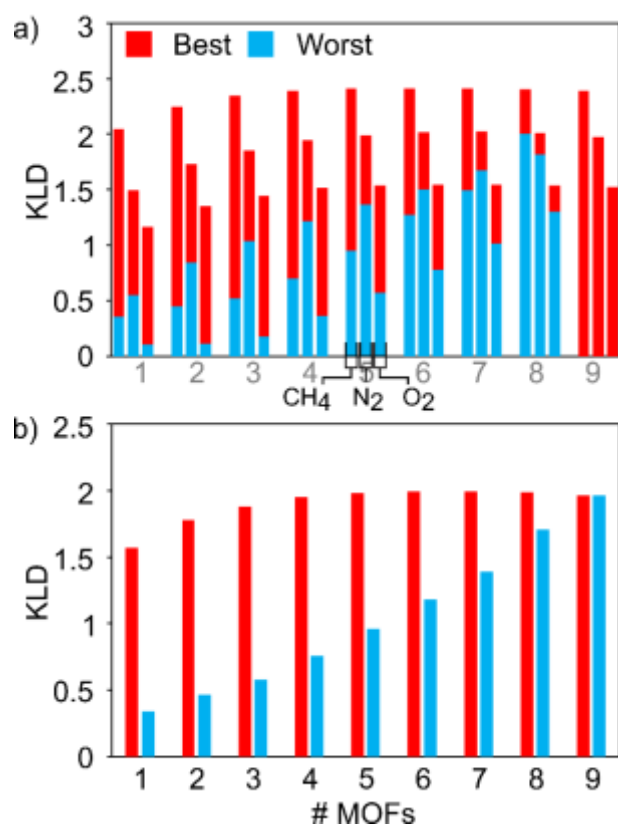


Figure 15. Averages over nine experiments: a) KLD vs number of MOFs in an array, predicting CH₄ (bar 1), N₂ (bar 2), and O₂ (bar 3). The best array KLD is read at the top of the red and the worst is the blue portion.

b) This is the same as a) but averages the KLD over all three gases at each array size.

For example, the *best* five-MOF array predicted CH₄ with an information content of 2.41, while the *worst* predicted CH₄ at an information content of 0.954. Furthermore, as hypothesized from the mass adsorption trends, the KLD values were highest for CH₄, followed by N₂, and lowest for O₂. In Figure 15b, we averaged the three component KLD values together for each array size, which allowed us to compare the best and worst arrays on their overall performance. Similar to the binary case, the best arrays peaked in KLD around five MOFs, at 1.98, and the worst arrays exhibited a steady increase as MOFs were added.

Distinct from the binary case, in the ternary analysis, the arrays that best predicted each component were not always constructed of the same MOFs. For example, the five-MOF array which best predicted CH_4 in experiment #1 was IRMOF-1, HKUST-1, UiO-66, ZIF-8, and MgMOF-75, and for O_2 it was IRMOF-1, HKUST-1, MgMOF-74, MOF-177, and NU-100. One advantage of this component-specific information is that we can tune a sensor array for a particular gas if necessary or optimize one to maximize detection of all components. Thus, if we are designing a sensor array for low concentration methane detection, we can construct an array which maximizes information for CH_4 prediction specifically.

Figure 16 depicts the abilities of the two best and two worst sensor arrays at predicting individual gas mixture components for experiment #1 as listed in Table 2 (i.e., 0.1 CH_4 , 0.75 N_2 , and 0.15 O_2). The plots contain probability curves for all three gas components, representing the complete predicted mixture from an array. Because the best and worst arrays differ between the three gases, a new KLD metric was used to rank the arrays per their performance over all components. For each array, the *product* of their KLD values for each gas was calculated and listed on each plot in Figure 16.

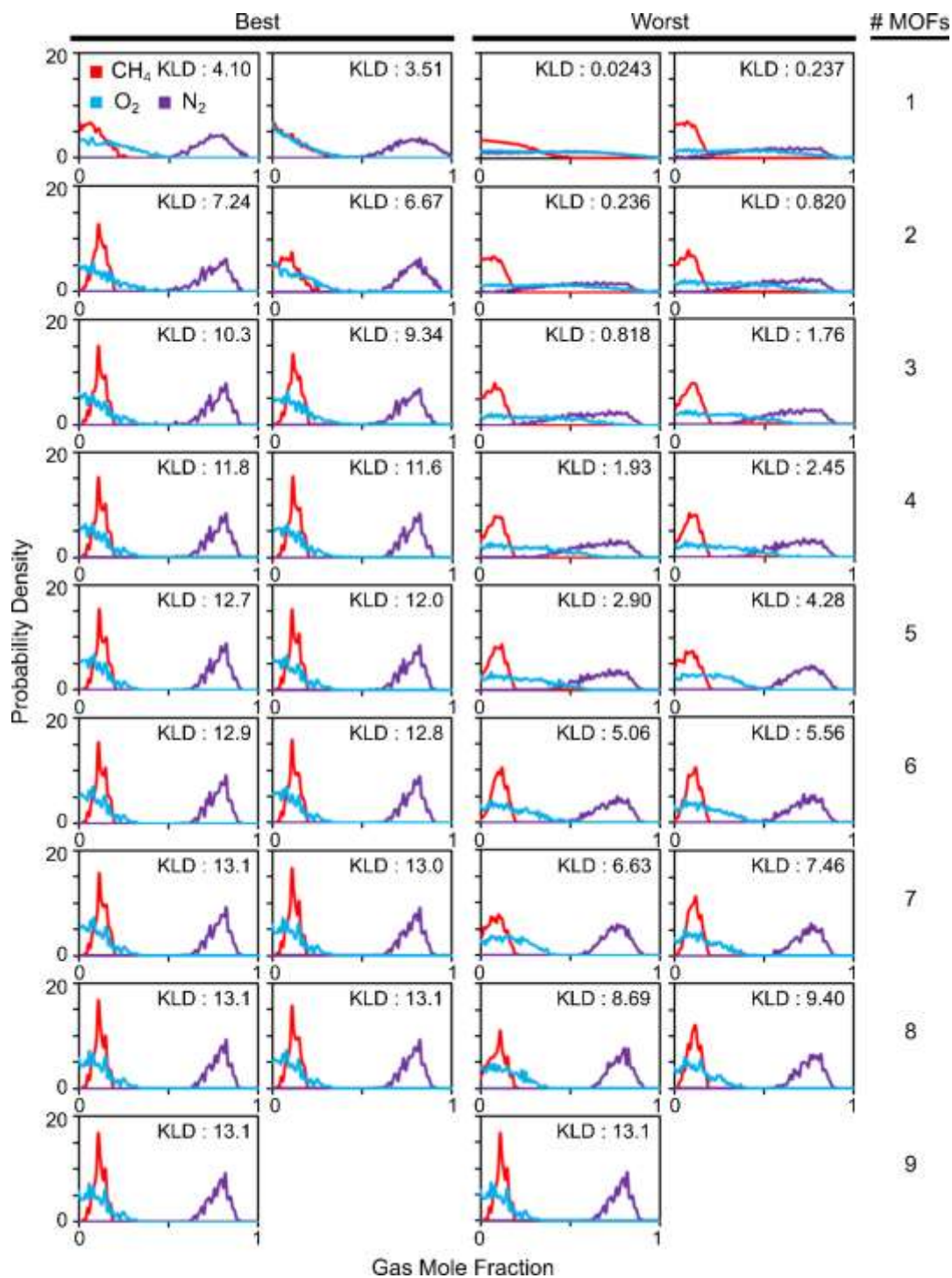


Figure 16. The probability versus concentration of CH₄, N₂, and O₂ is plotted for the top and bottom two arrays, showing increasing confidence through narrowing peaks as the array size increases. Arrays are ranked by the product of the KLD values as they relate to each gas.

We see the probability distributions in Figure 16 narrow with increasing array size, allowing for clearer resolution of the predicted gas mixture concentration. In agreement with the KLD trends (see Figure 15), CH₄ and N₂ are better resolved than O₂, with sharper probability peaks. Even with nine MOFs, the mole fraction of O₂ in the mixture is not as well resolved; however, if the other gas concentrations are known, then the O₂ concentration is such that the sum of all mole fractions is one.

In Figure 17, we take a closer look at the nine-MOF array plot from Figure 16, where the predicted concentrations are shown more precisely.

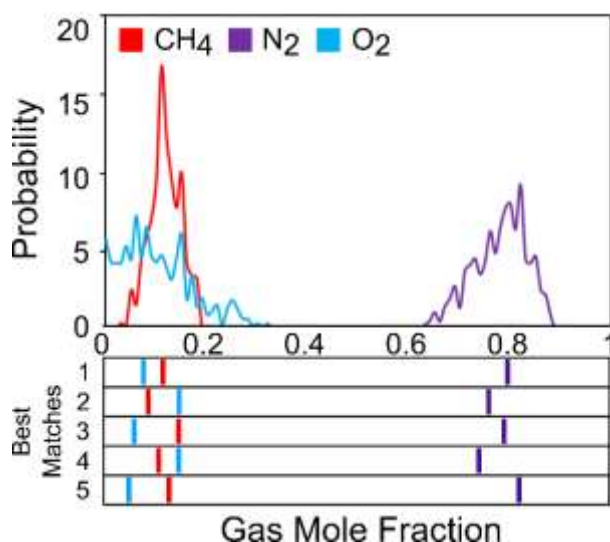


Figure 17. The nine-MOF probability vs concentration plot from Figure 16 above. Below are the five highest probability gas mixtures for experiment #1 in Table 2, whose concentrations align with the peaks in the plot.

The peak probabilities for each gas occur at 0.1 CH₄, 0.8 N₂, and 0.08 O₂, whereas the experiment values are 0.1 CH₄, 0.75 N₂, and 0.15 O₂. Note that adding the mole fractions associated with each peak does not necessarily have to add to one. However, by looking at specific

gas mixtures in the simulation data set, we can list the best matches, for which the mole fractions of the gas components do add to one. In Figure 17, one can see that the top five best matches have each of their gas component concentrations near the peaks of their corresponding probability distributions.

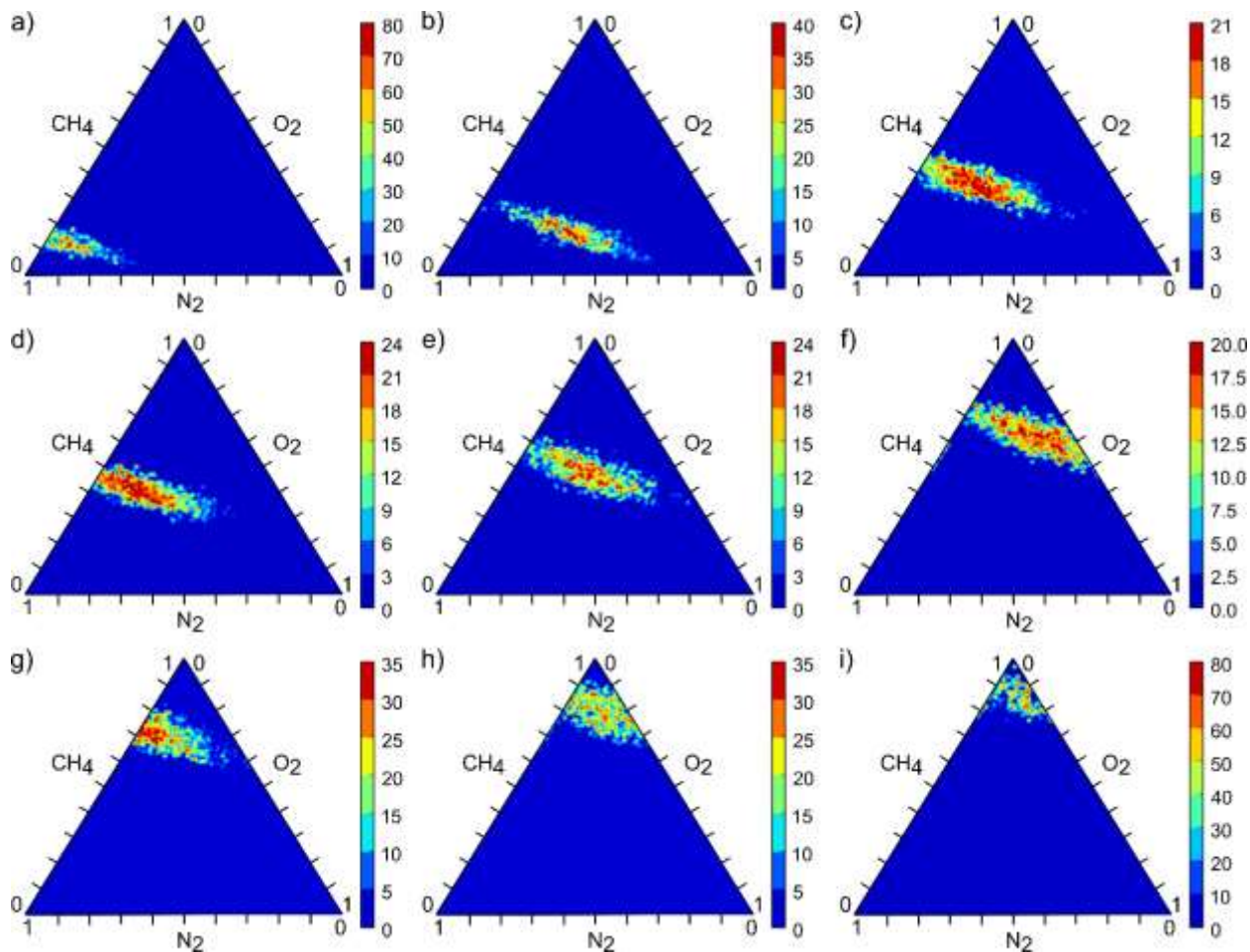


Figure 18. Probability vs concentration for all simulated gas mixtures in the nine-MOF array, a) – i) being experiments 1–9.

Although similar KLD values were found for the nine-MOF array and best four-MOF array (Figure 15b), which might lead one to think they have very similar performances in resolving gas

mixtures, we found that even small differences in KLD values can have significant impacts on their associated probability distributions. Looking at Figure 18 and Figure 19, we compare the predicted compositions for each experiment for the nine-MOF array and the best four-MOF arrays, respectively.

Note that, in Figure 18 and Figure 19, best and worst is defined according to KLD values for CH_4 probability distributions only. Across all nine experiments, the gas mixtures are more sharply resolved for the nine MOF array (Figure 18) than for the best four MOF arrays (Figure 19). Thus, the nine-MOF array is clearly better able to narrow down the concentrations of each component, even though its KLD values is only marginally higher than for the best four-MOF arrays.

We see also, in Figure 19, an example of how the probability distributions are affected by the choice of gas component. Since the data in Figure 19 are for the best four-MOF arrays, according to KLD values for CH_4 , there is a narrower range for possible CH_4 mole fractions, whereas for O_2 and N_2 , the ranges are much larger. The red points, which correspond to the highest probabilities, generally cover a greater range of possible N_2 mole fractions, so the precise concentration of N_2 is more difficult to ascertain (at least for the MOFs considered in this study). Similarly, for O_2 , the predicted concentrations are spread over a large range in most experiments. If we were interested, instead, in the concentration of one of the other components, then we can choose to maximize arrays according to the appropriate KLD values. When it is necessary to resolve all mixture components, this KLD ranking system can facilitate making informed trade-off decisions, where greater precision in resolving one component comes at the cost of less precision in detecting another.

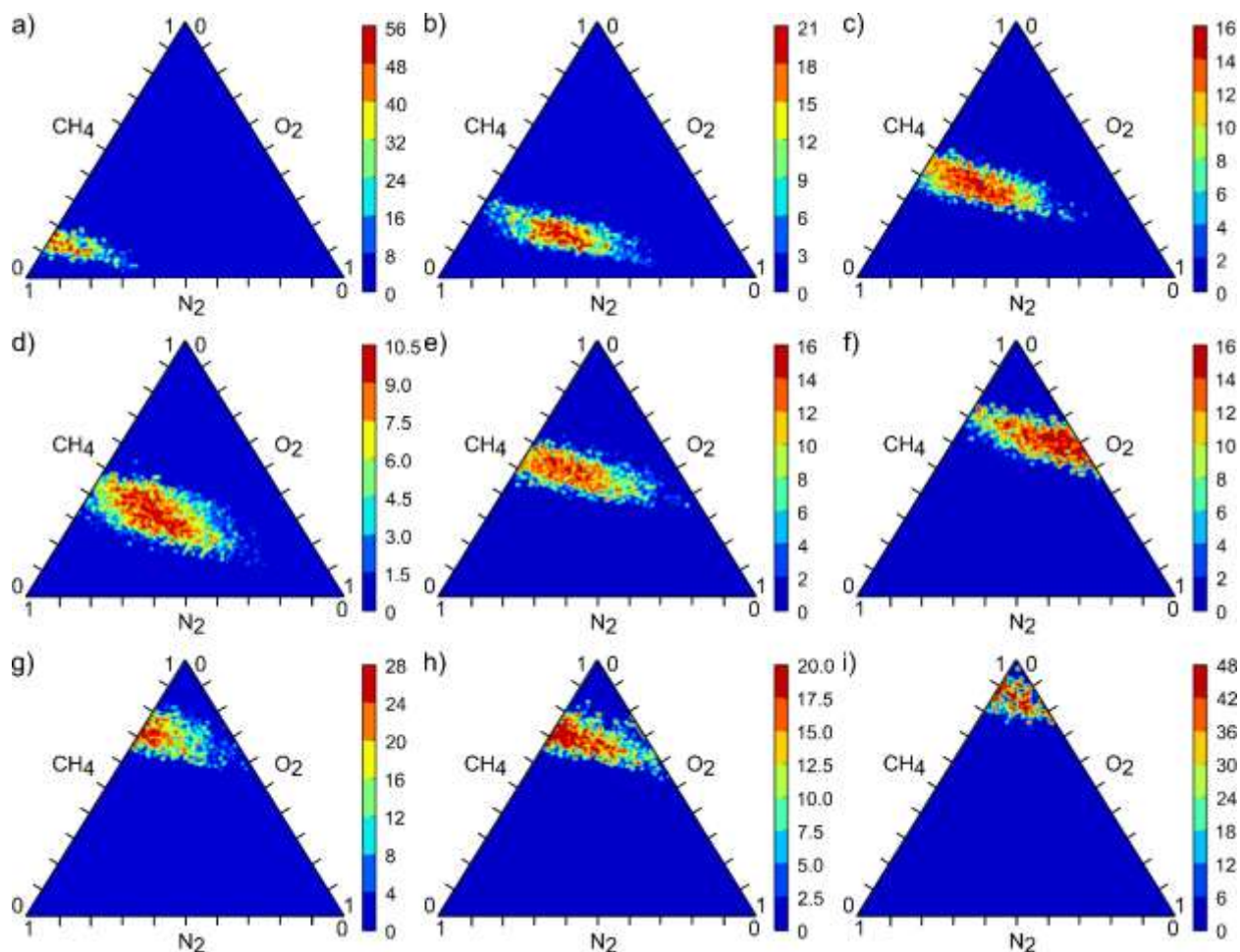


Figure 19. Probability vs concentration for all simulated gas mixtures for the best four-MOF arrays a) – i) being experiments 1–9.

3.7 Conclusion

This work specifies a method for screening arrays of MOFs via atomistic molecular simulations and ranking them based on concepts from information theory. We applied this method to the full range of gas mixtures composed of CH₄, N₂, and O₂, where individual MOF sensing elements were assumed to only detect total mass adsorption. When considering all of the four-

MOF arrays, from a library of nine possible MOFs, of which there were 126 possible configurations, the *best* ones produced significantly more information than the *worst* ones, highlighting the benefit of computational array optimization over a trial-and-error approach. In general, arrays were able to resolve CH₄ concentrations more precisely than for N₂ and O₂. However, if precise detection of either N₂ or O₂ was a desired feature, arrays could be optimized to produce the highest information content for those specific gas components. Alternatively, our method may be used to optimize an array to maximize sensitivity for trace gas detection, which is of critical importance for detection of methane leaks in air. Furthermore, we will consider additional components present in air, including CO₂ and water vapor, as their presence will have an impact on MOF behavior. In the next aim, we consider larger arrays, and larger libraries of MOFs, to explore how much further we can optimize the design of MOF-based sensor arrays.

4.0 Intelligent Selection of Metal-Organic Framework Arrays for Methane Sensing via Genetic Algorithm

Jenna A. Gustafson, Christopher E. Wilmer

In progress, to be submitted to ACS Sensors

4.1 Motivation

This study is a progression of our previous work, to be applied in conjunction with the aforementioned processing and optimization tools. In the prior section, we established a method that allows for 1) the prediction of an array's output response and 2) the optimization of array construction. Our previous studies, with the intention of establishing working methods, only used up to nine different MOFs in our analysis. In practice, there are significantly more sensing materials from which to choose and more gases for complex sensing applications.

Initially, we used a brute force approach, comparing the performance of MOF arrays by testing every possible array combination. This was feasible for nine MOFs, where there were 511 potential arrays. However, the amount of combinations increases exponentially upon increasing the possible number of elements, as demonstrated in Figure 20.

As it is impractical to test hundreds of sensor arrays experimentally, it is also infeasible to computationally perform brute force simulations of trillions of possible arrays. Scaling linearly from the time to calculate 511 arrays in our prior work, this would take 9,308 years. Therefore, in this section, we implement a search algorithm for the large-scale intelligent selection of arrays.

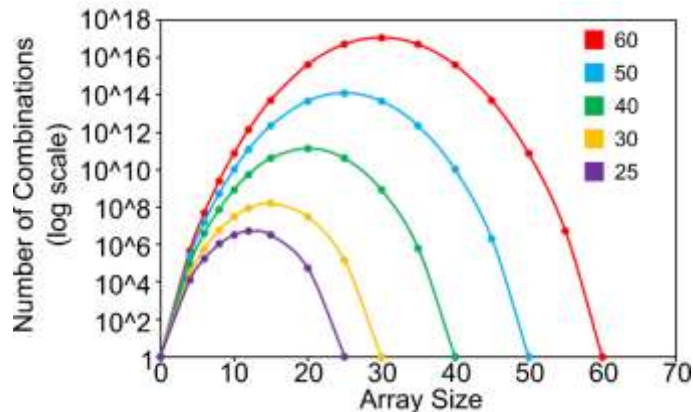


Figure 20. Possible combinations on a log scale vs array size, varying the possible number of materials to choose from (60, 50, 40, 30, 25).

Various types of evolutionary algorithms have been used for the screening of hypothetical materials, including MOFs and functionalized materials, constructed with many potential building blocks. More generally, genetic algorithms have been used for search optimization for many years [101,102]. A common approach to efficiently solve such problems is to find the closest generic optimization problem and adapt its best-known algorithmic solution. Our problem closely compares to the *knapsack problem*, a common combinatorial optimization problem [103]. Specifically, the knapsack problem is to select a group of items that do not exceed the maximum weight while maximizing the total value, given a set of materials with weights and values [103]. Our challenge is similar, to construct an array of MOFs that maximizes the overall information content from the sensor's output signal. Further, we simplify this task by constraining our search to select MOF arrays of a specified size, rather than search for the overall best performing array. Since this is not a traditional knapsack problem, we must adapt existing algorithms to optimize arrays with no obvious best approach. Consequently, we will attempt solutions using a genetic algorithm (GA), to solve the knapsack problem, adapted to our specific case [104–106].

4.2 Genetic Algorithm

The general approach to genetic algorithms has been well documented but can be described here simply. Please refer to Figure 21 below, following each step in the process.

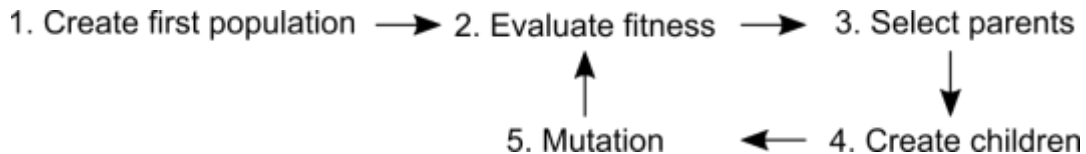


Figure 21. Flow diagram of steps in a genetic algorithm.

First, a challenge with genetic algorithms is the encoding of a MOF array as a genome (i.e., a string of integers), for which there is not one correct solution. GA encoding is how each array is represented as code, as in the example shown in Figure 22 below.

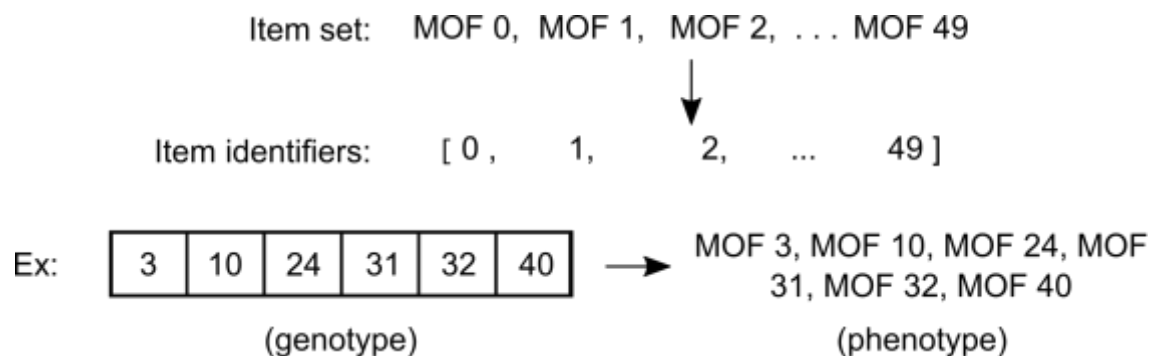


Figure 22. Example of an encoded MOF array in a genetic algorithm.

Analogous to human genetics, the human genotype (DNA) is the GA encoding and the phenotype (physical features) is the solution expression of the GA. The genotype is often defined as a string of zeros and ones, to represent if an item is expressed in that individual, as with DNA

identifiers. Then, the phenotype is the qualitative collection of expressed items, whether it is a string of letters or a group of chosen materials, as with a person’s physical traits.

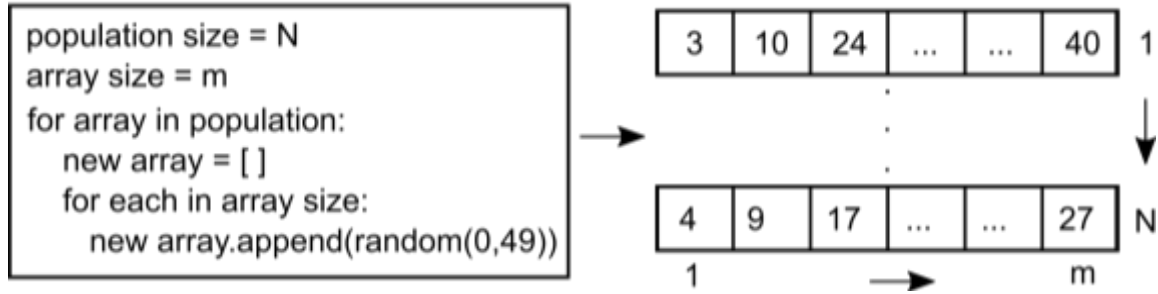


Figure 23. Procedure for creating the first population of arrays.

Specifically, our solution requires selecting candidate arrays of MOFs, where we impose the limitation that each array must be a specified size. We encode each array as a list of numbers, where each index value corresponds to the list of all MOFs, which is then expressed as the list of MOF names selected (Figure 22). With this method, we still encounter the potential of repeat MOFs in an array, which we have mitigated by implementing checks that do not allow solutions with duplicates. Initially, we create a random first population of possible solutions, where the process is shown in Figure 23.

If we have a population size N and an array size of m , for N arrays we select random numbers from 0 to 49 m times to make up each array. The result is a set of N lists of length m , where we impose the constraint of no duplicates in one array.

The next step is to evaluate the “fitness” of each array solution, defined here as a product of three KLD values from a candidate array (Equations 4-1 and 4-2). Each KLD value is the result of an array’s probabilistic prediction of one gas in a mixture; thus, since we are testing three component mixtures there are three KLD values for each array.

$$KLD = \sum_i^N P_i \log \frac{P_i}{Q_i} \quad (4-1)$$

$$Fitness = KLD_{CH_4} \times KLD_{O_2} \times KLD_{N_2} \quad (4-2)$$

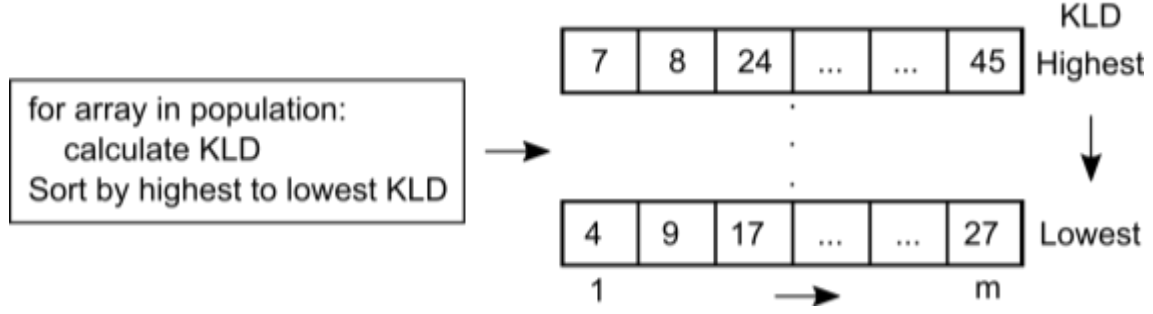


Figure 24. Fitness (KLD) is calculated for each MOF array and ranked from highest to lowest.

The probability at each mole fraction is represented by P_i , and a reference probability of Q_i is a probability equivalent to $1/N$ (i.e., a uniform probability distribution), where N is the number of points (mole fractions) we have from 0 to 1 for each gas. We then rank the arrays from highest to lowest KLD, as in Figure 24. Please see Appendix E for further details.

The array with the highest KLD is saved as the current best result, and the next generation of candidate arrays is created. We then select parents from the ranked solutions to be used in creating the next generation of arrays, as shown in Figure 25. The solutions with the highest fitness are kept as parents, in our case the top 20% of all arrays. Additionally, we randomly select 20% of the population as parents so that we still have genetic diversity and avoid local optima. Next, the parents are combined to create the next generation, “children”, in a process called *crossover* (Figure 26). As with human DNA, a child inherits pieces from both parents, so with each bit in a

solution we randomly select which parent to take from, resulting in some combination of the parents' "DNA". Equal probability is given to each parent.

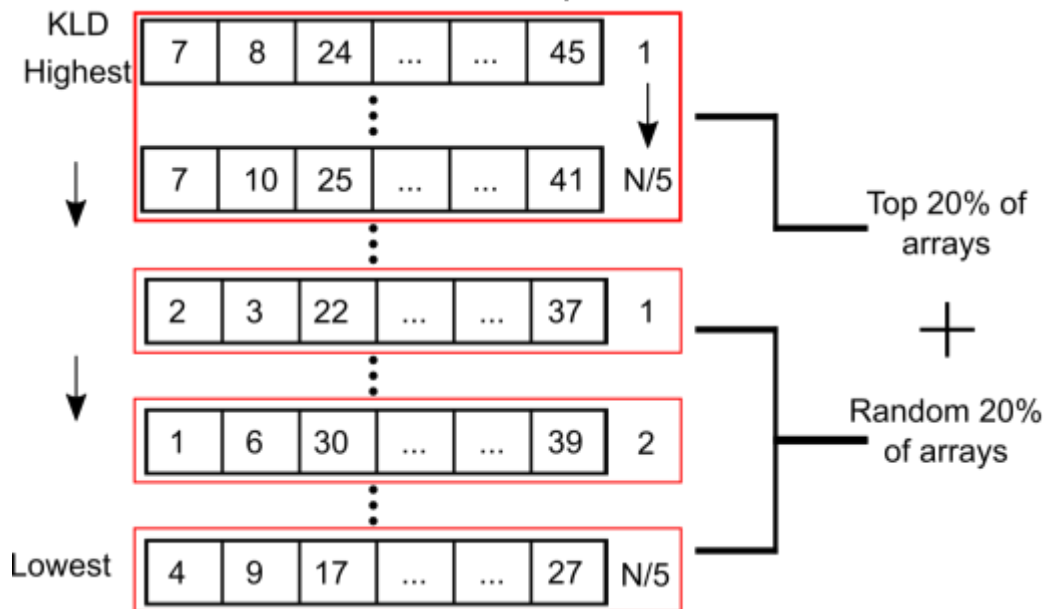


Figure 25. Selecting parents from the top 20% of ranked arrays and random 20% of arrays.

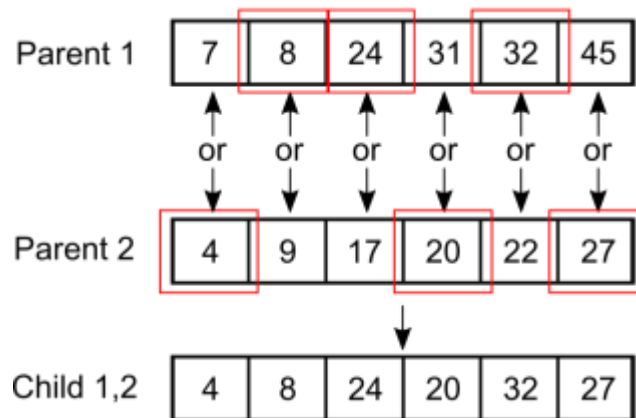


Figure 26. Crossover of two parents to create a child.

Our approach is to randomly retain each array element from one parent or the other, giving equal probability to each parent. The result is a new array with random elements from each parent. Combinations of parents are taken until enough children are created to make up the next generation. During the crossover process, each array is again checked for duplicates and if found they are replaced with a different random MOF. Once the next generation of children has been created, each may undergo a process called *mutation* (Figure 27). This is a user-defined probability for an element in to be mutated, often set to a low value, and we have chosen 0.1% mutation for the following analysis, unless otherwise specified. In our case, if an element is selected for mutation then a random number (from 0 to 49) is selected to replace it, then the array is checked for duplicates (as in crossover). Mutation allows for genetic variation and is used to avoid local optima, but the optimal mutation rate itself is found by trial-and-error.

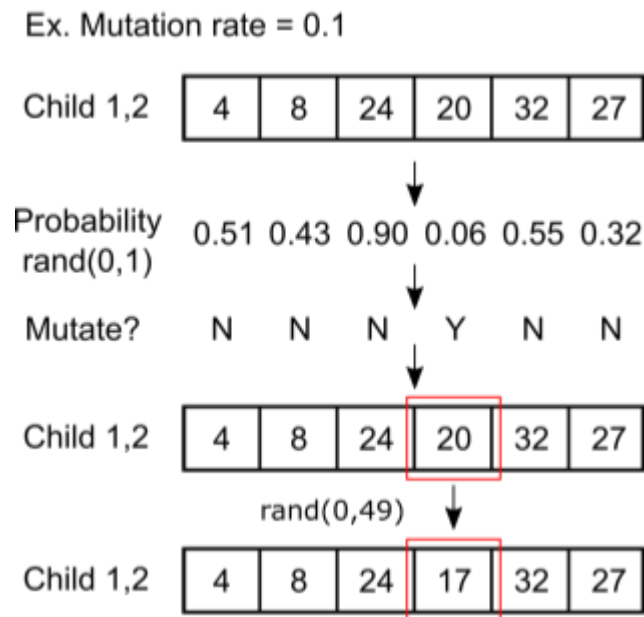


Figure 27. Mutation of an individual, with a mutation rate of 0.1

After mutation, the fitness for each solution in the new population is evaluated, and the above process repeats for a user specified number of generations. After each generation is evaluated, we save the top performing result, which we expect to yield the best possible array.

Genetic algorithm optimization is non-trivial, where parameters of population size, mutation rate, and number of generations may be adjusted with differing results. The mutation strength may have a great influence on algorithm performance, as there should be some genetic variation, but a high rate may lead to randomness. We are posed with the task of optimizing parameter values to yield the most efficient solution. We have built a genetic algorithm to select the best performing arrays for any size, where we can predict the best MOF arrays on a time scale of minutes. Below, we describe the process of optimizing our algorithm for selecting the best arrays from 50 possible MOF structures.

4.3 Predicting Gas Adsorption in MOFs

In a continuation of our previous work, we are designing MOF arrays for methane-in-air sensing, using simplified mixtures of methane, oxygen, and nitrogen. For the sake of efficiently establishing an optimization methodology, we are not considering additional components found in air such as water, carbon dioxide, etc. Simulations become more complex as more gases are included in the mixtures. We performed grand canonical Monte Carlo (GCMC) simulations using the software package RASPA, [96] calculating adsorption for all mixtures of methane, oxygen, and nitrogen, varying the concentration of each gas by 2%. The result was the total mass uptake of 1,326 mixtures in 50 MOF structures at 298K and 1 bar. MOF structures were chosen to

represent a range of surface areas and were obtained via the CoRE MOF Database.[\[107\]](#) Please see Appendix A for a list of structures and further simulation details.

4.4 MOF Array Sensing

Previously, we demonstrated how gas mixture compositions could be predicted using only total mass adsorbed onto each MOF in an array.[\[108\]](#) The mass change of an individual MOF rules out certain gas mixtures, but leaves many others as likely, which can be described by a probability distribution. We found in earlier work that by considering input from multiple MOFs via joint probability distributions, the range of likely gas mixtures narrows substantially. Then, the performance of each MOF array could be quantified using a metric called the Kullback-Liebler Divergence (KLD), which measures the information content of a probability distribution. In *this* work we calculate KLD values for many arrays of various sizes that can be constructed from a library of 50 different MOFs. Even though calculating a KLD value is relatively inexpensive computationally (particularly when compared to simulating gas adsorption), there are trillions of array combinations (see Figure 20) and so brute force screening cannot be used.

Using probability distributions of gas species and KLD values to quantify the performance of sensory arrays differs from other commonly used methods to identify gases from electronic nose data, which are typically based on fingerprints, training data, and pattern recognition.[\[10,12,26,94\]](#) Electronic nose performance is often scored based on a sensor's percentage of correct classifications of compounds.[\[26,94\]](#) Common classification techniques include principle component analysis (PCA) or partial least squares-discriminant analysis (PLS-DA) with leave-one-out training, which reduce data dimensionality and group responses.[\[82\]](#) In contrast, our

approach uses a molecular model of gas mixture adsorption instead of fingerprints to convert sensor signals into detailed gas composition information.

4.5 Genetic Algorithm Analysis

4.5.1 Comparison to Brute Force Analysis

Initially, we compare the genetic algorithm results to a known *correct* result to validate its accuracy for predicting the best MOF arrays (which is only possible for small array sizes). Therefore, we evaluate one, two, three, and four MOF arrays (from a library of 50 MOFs) for which we can easily compute KLD values for all possible combinations (See Figure 28). We see that a small percentage of arrays have significantly better performance than the rest of the set. Please note that all reported KLD values are in units of bits³ throughout the study.

Subsequently, we compare the top MOF arrays from our genetic algorithm to the known correct results, as shown above in Figure 28. The GA was run for array sizes of one, two, three, and four, where each run included fifty generations with a population size of one hundred and mutation rate of 0.1%. The black points in Figure 28 represent the best array found in each of five GA runs for each array size. We see that algorithm clearly performs well having found the global best array for each array size, in some cases in all five runs. Specifically, the GA found the global best array 3 out of 5 times for arrays of one and four MOFs, and 5 out of 5 times for arrays of two and three MOFs. This gives reasonable confidence that the GA will yield high-performing arrays when considering larger sizes as well.

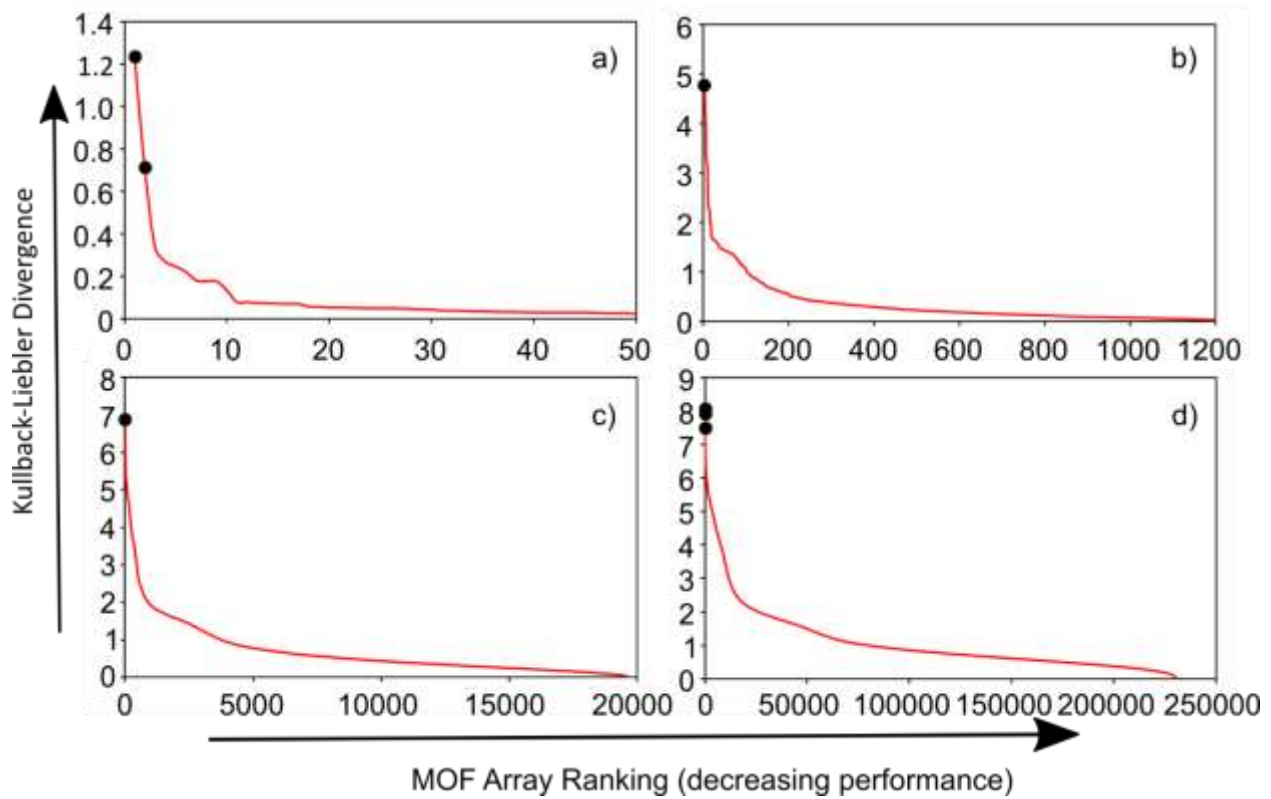


Figure 28. KLD (in bits³) vs ranking of every MOF array for size of a) 1 MOF, b) 2 MOFs, c) 3 MOFs, and d) 4 MOFs. Black points represent best arrays found from the GA over five runs, where some are overlapping.

We can display the improvement of arrays' abilities to predict gas mixture compositions by showing probability versus mole fraction relative to each gas, as the array performance decreases, from one to four MOF arrays (Figure 29). In this study, the unknown mixture we want to predict is 10% methane, 20% oxygen, and 70% nitrogen. The probability plots correspond to arrays evenly spaced from best to worst along the analogous plots in Figure 28 above. The trends in probability validate the KLD ranking metric as a tool for informing the selection of sensing materials relative to electronic nose design.

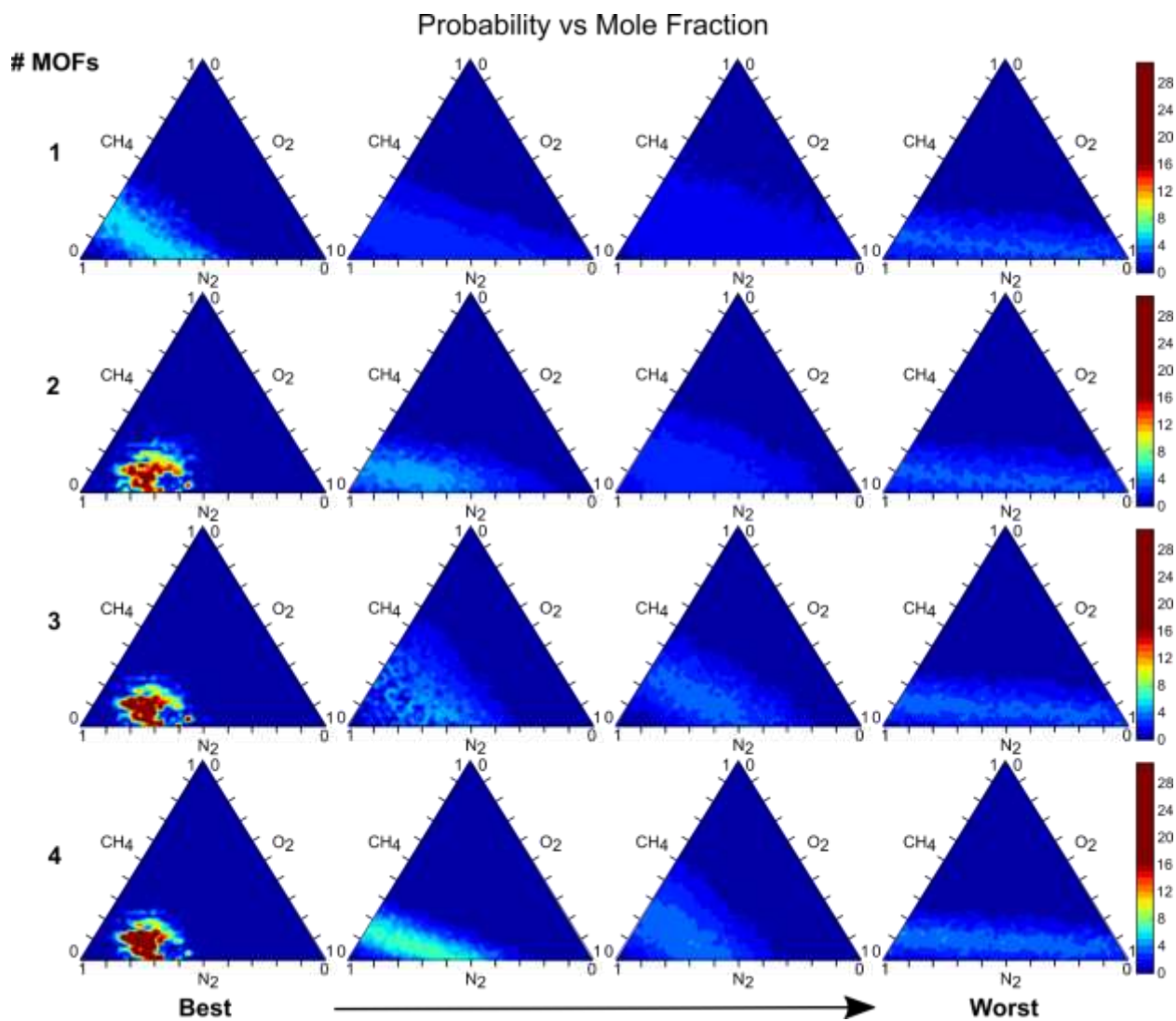


Figure 29. Probability density versus mole fraction of CH₄, O₂, and N₂ for array sizes of 1, 2, 3, and 4 for predicting 0.1 CH₄, 0.2 O₂, and 0.7 N₂. Results are from the best arrays all on the left, and subsequent arrays are evenly spaced based on decreasing performance (100th, 66th, 33rd, 0th percentile, ranking from left to right).

4.6 Genetic Algorithm Parameter Analysis

4.6.1 Varying Population Size

In the interest of optimizing the GA parameters to maximize computational efficiency, we investigated the trade-off between the algorithm's convergence time and performance by varying the population size.

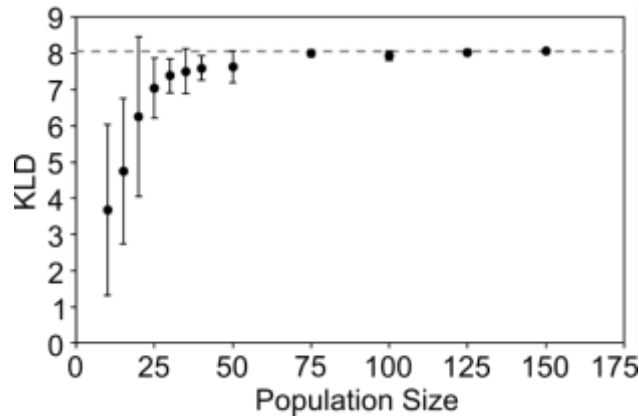


Figure 30. Plot of KLD (in bits³, averaged over 10 GA runs) versus population size, for selecting the best 4 MOF arrays. Dashed line is the KLD of the best possible 4 MOF array.

Displayed in Figure 30 is a plot of the KLD value, averaged over multiple GA runs, against population size, for predicting the best 4 MOF arrays, using a 0.1% mutation rate and 50 generations. The KLD steadily increases with population size from 10 up to around 40 arrays, then maintains approximately the same value for populations up to 150. Additionally, error bars indicate the deviation of the KLD from the average, showing less deviation in the GA result between runs with larger population sizes, therefore indicating an improvement in the GA reliability. The dashed line represents the best 4 MOF array, at a KLD of 8.062 bits³.

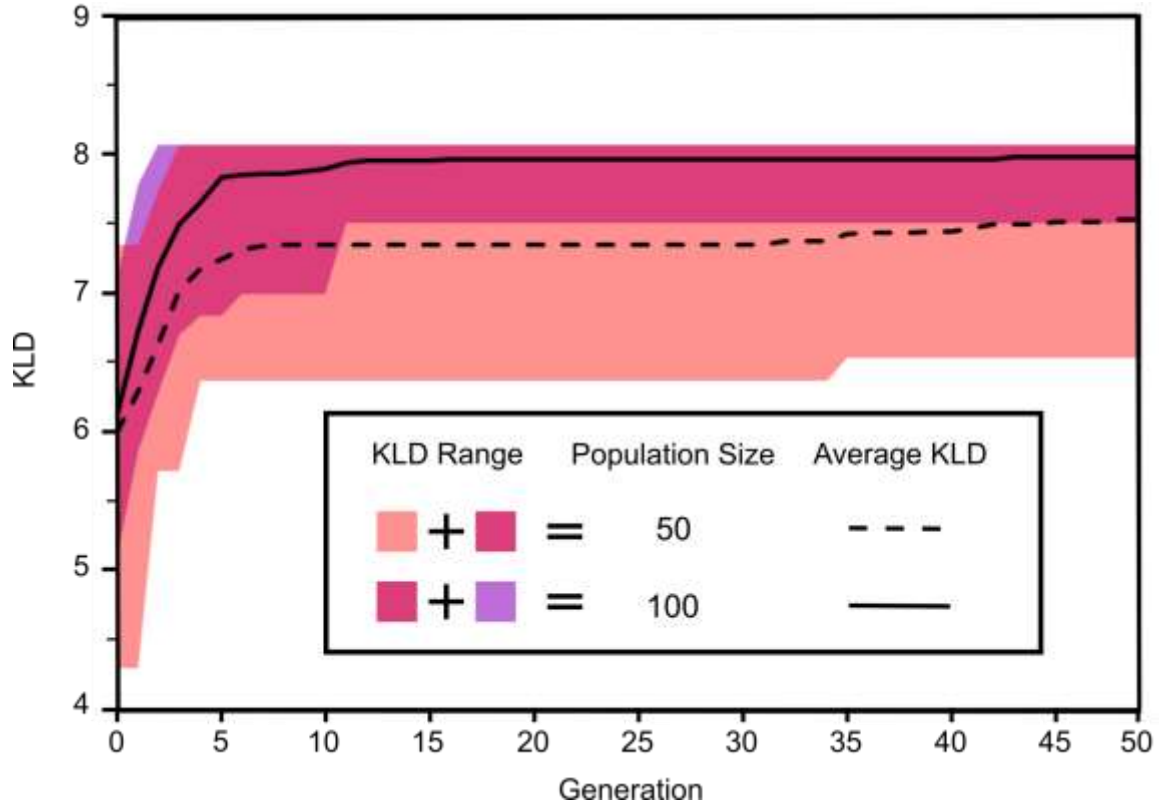


Figure 31. Plot of average KLD (in bits³) with the range of possible values shaded versus the generation for selecting 4 MOF arrays with a population size of 50 and 100 arrays. The upper bound peaks at the KLD for the best possible 4 MOF array.

Further, we can show the range of KLD values obtained over the range of one GA run, averaged over multiple runs, and plotted versus generation (Figure 31). We compare population sizes of 50 and 100, where the larger population again yields a consistently high-performing MOF array prediction. The shaded region is the range of KLD values obtained at each generation, where the population size of 100 shows little variation compared to 50.

From this, we see that the algorithm converges well before 50 generations for a population size of 100, where the best array is consistently obtained around generation 13. Thus, to improve

efficiency in practical implementation, the number of generations may be lowered to reduce algorithm time.

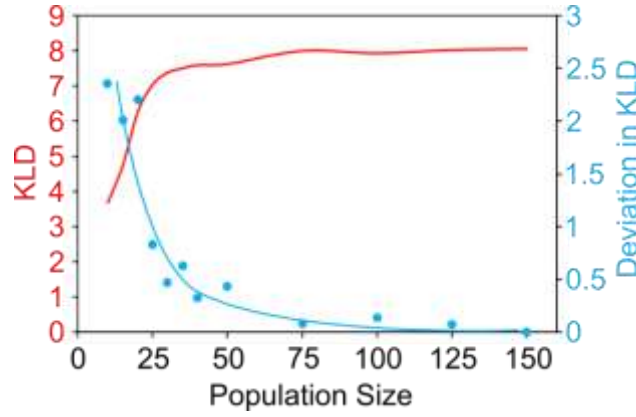


Figure 32. Plot of normalized values for KLD (in bits³) and deviation in KLD versus the population size, for selecting 4 MOF arrays.

Finally, the ideal population size would minimize the variation in KLD values for a *high* KLD, while also minimizing the algorithm time. Moreover, minimizing the variation will give more confidence in each GA result, so one GA run will be enough to predict the best array. In Figure 32, we look at these metrics on the same plot, where the values of KLD and deviation from average KLD have been normalized relative to their highest possible values. Assuming reliability (i.e. low KLD variation across runs) is important, we determine the optimal population is the smallest at which we obtain the maximum KLD values with little deviation. Thus, we have confidence in using a population of 75 in the full algorithm implementation.

4.6.2 Varying Mutation Strength

Previously, we could predict that increasing population size would improve the GA outcomes; however, the same intuition does not hold for the mutation rate. Using a population size of 75, for 50 generations, a variety of mutation rates were tested for array sizes of 4, 10, and 25 MOFs. Because it is a percentage, mutation rate scales with array size (i.e. more elements are subject to change as array size increases), so a range of sizes were tested to select an optimal rate for any array size. In Figure 33 we show the average KLD values normalized relative to the highest KLD found for that size array versus the mutation rate for a range from 0% to 20%.

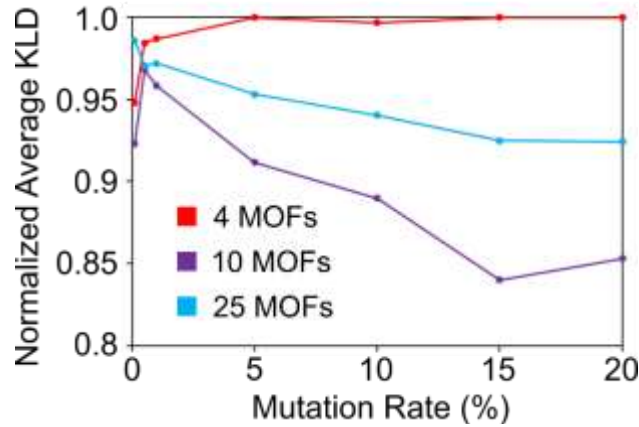


Figure 33. Plot of KLD values normalized, relative to the highest KLD for that array size, versus mutation rate in percent for 4, 10, and 25 MOF arrays.

From this, it is evident that array size affects the mutation rate performance. Mutation rates above 5% yield good results for a 4 MOF array, but for 10 and 25 MOFs the algorithm performance steadily drops as the rate increases. Interestingly, the trends don't show that mutation has a greater effect as array size increases; therefore, we cannot extrapolate the performance to larger array

sizes. Regardless, the overall best mutation rate is in the range of 0.5 to 1%, so we take a closer look at values in this region (Figure 34).

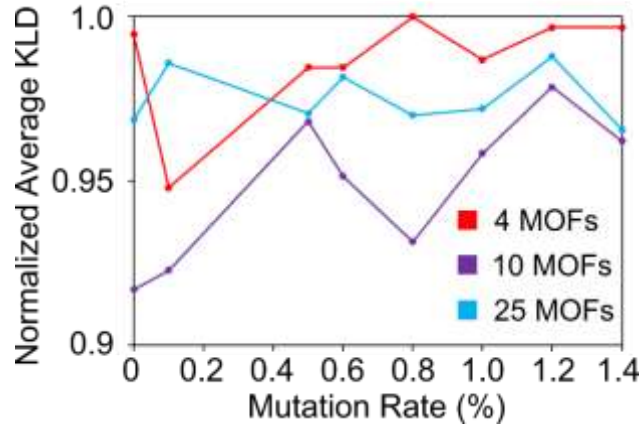


Figure 34. Plot of KLD values normalized, relative to the highest KLD for that array size, versus mutation rate, 0 to 1.4 percent, for 4, 10, and 25 MOF arrays. (Please note the difference in the y axes values from Figure 10.)

In the range of 0.5% to 1.4% mutation rate, there is a consistently high average KLD value for all array sizes tested, with no strong trends. Looking at Figure 34 above we can safely choose a mutation rate in the range of 0.8% to 1.2% and we will achieve comparable results. Although we do not know the correct results for array sizes larger than four, we still have confidence in this range of mutation rates, because in the literature[[102](#),[109](#)] mutation rates of around 1% are prevalent.

4.6.3 Increasing Array Sizes

Finally, we report results MOF array sizes ranging from one to fifty, where as we have previously reported,[[95](#),[108](#)] the KLD values increase with increasing array size (Figure 35). Although we cannot know the global maximum KLD for each array size, given the MOFs in our study, we are confident that we are finding at least a comparable local maximum from the GA. If we extended the study to additional MOFs, we theorize the KLD will continue to increase linearly until it reaches the maximum around 180 bits³, where it will level off. Please see Appendix E for further details.

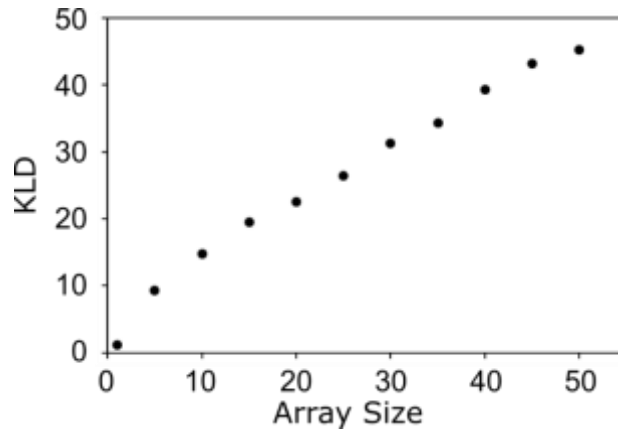


Figure 35. Plot of *best* KLD value (in bits³) versus array size, using a population size of 100 and a mutation rate of 0.1%.

Lastly, we plot the results from each best predicted array for array sizes of five through fifty MOFs, as in Figure 36. The plots represent the probability versus concentration of all gas mixtures of methane, nitrogen, and oxygen, where the highest probability point is associated with a test gas mixture of 0.1 CH₄, 0.2 O₂, and 0.7 N₂. As the array sizes increases, the probability also

increases in a concentrated area, with the results steadily improving up to fifty arrays. The probability results support the assumption that the GA accurately produces the best arrays, in that we achieve a precise prediction of the unknown gas mixture. At an array size of fifty we see the highest probability at one gas mixture, rather than a cluster of mixtures.

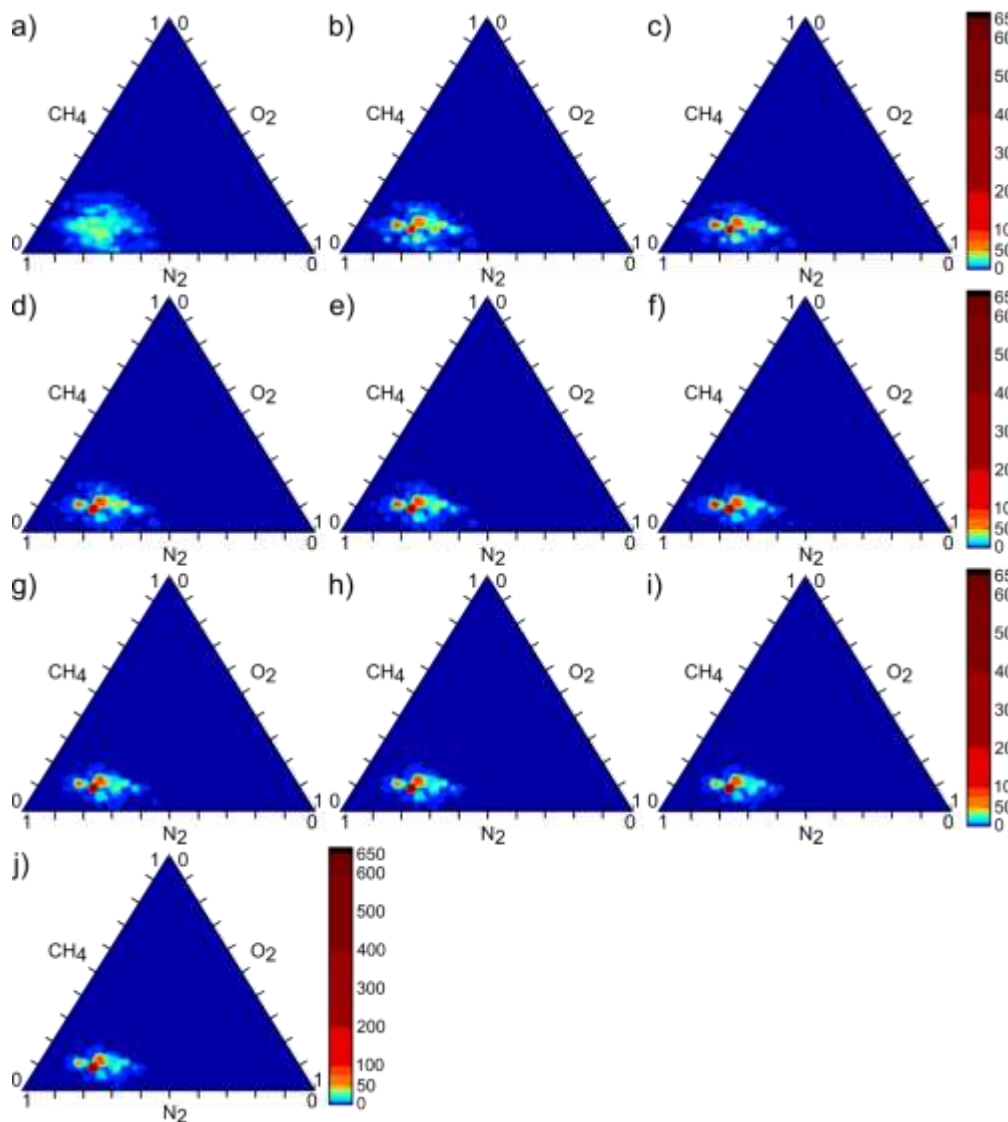


Figure 36. Plots of probability versus mole fraction of CH₄, O₂, and N₂ for the best predicted arrays of a) 5, b) 10, c) 15, d) 20, e) 25, f) 30, g) 35, h) 40, i) 45, and j) 50 MOFs.

4.7 Conclusions

Despite the existence of MOFs for over 20 years, limited attention has been given to their gas sensing capabilities. In a continuation from our previous work, we have successfully ranked arrays of MOFs for methane detection in the presence of oxygen and nitrogen. As more MOFs are considered, their optimal selection becomes an exponentially more computationally expensive problem. We have implemented a genetic algorithm to efficiently search for the best array configurations, testing a total of 50 MOF structures with 1,326 gas mixtures of methane, oxygen, and nitrogen.

After comparing different GA parameters for a range of array sizes, we concluded that a population of 75 arrays and a mutation rate of around 1% yields the best results, regardless of array size. Additionally, it was found that the average KLD values for the best arrays increase as MOFs are added to the array, all the way up to 50 MOFs. Importantly, we have shown our ability to find the optimal MOF arrays in minutes, even for large array sizes, as opposed to thousands of years doing an exhaustive search. We hope that these tools may be used to inform experimental design of MOF sensors going forward.

In future work it will be necessary to include all appropriate MOFs from the CoreMOF database. Depending on the pool of MOFs from which we can choose, there may be ten other MOFs which perform better than all fifty from this study. Thus, it is important to include all potential materials to obtain the best possible arrays in order to tackle today's gas sensing challenges.

5.0 Development of Metal-Organic Framework (MOF) Coated Bulk Acoustic Sensors for Carbon Dioxide and Methane Detection

Jenna A. Gustafson, Jagannath Devkota, Christopher E. Wilmer, Paul R. Ohodnicki

Submitted to Sens. Actuators B Chem.

5.1 Introduction and Motivation

Acoustic gas sensors have been used for decades, serving a range of applications from food quality assessment to environmental monitoring [4]. While such devices work well for industrial purposes, it is a continuing challenge to reliably detect complex gas mixtures at the ppm or ppb levels. Many current gas detectors use polymer films or metal-oxide surfaces as sensing materials, where it has been a challenge to achieve the selectivity and sensitivity required to further improve sensing capabilities. However, newer classes of porous materials are being investigated to face these challenges, including zeolites [59–61], carbon nanotubes [29,54,55,110,111], and metal-organic frameworks (MOFs) [56,57,62,66,71,75,77,112]. Specifically, MOFs have been around for two decades, but limited exploration has been done regarding their sensing capabilities. MOFs are ideal potential candidates for mass-based sensing; they have high surface area, crystalline structures, which lend themselves to record-high gas adsorption and reproducibility during synthesis [56]. Additionally, there are thousands of known MOF structures, with a range of pore sizes and chemistries, allowing for tunability to target specific gases [77,107].

Further, the crystalline nature of MOFs lends itself to the prediction of their behavior in complex mixture environments using molecular simulations. Previous studies have established the validity of methods such as grand canonical Monte Carlo (GCMC) simulations for modeling adsorption of gases in MOFs, as they relate to sensors [68,77]. As previously stated, MOFs are of interest for mass-based sensing platforms, including surface acoustic wave (SAW) devices and quartz crystal microbalances (QCMs). As gases adhere to the surface of the MOF sensing layer, the propagating wave along the device is disrupted, resulting in a frequency shift of the mechanical wave. This frequency shift directly correlates with the total mass change of the MOF due to its interaction with the surrounding gas [25]. Therefore, predicting gas adsorption in MOFs is a viable approach to modeling the sensor signal from a SAW and/or QCM device.

Work has been done previously to fabricate MOF-based sensors by Devkota et al., showing their compatibility and reliability with both surface acoustic wave (SAW) and quartz crystal microbalance (QCM) devices [67]. QCMs are suitable candidate devices for electronic nose development, particularly due to their fast response times and portability [9]. In this study, we coat a QCM device with a sensing layer of the MOF called ZIF-8 [87], from which total frequency shifts are obtained as a result of gas mixture adsorption. Thus, we compare experimental and simulation results for a ZIF-8 coated sensor detecting various mixtures of CO₂, CH₄, and N₂.

5.2 Methodology

5.2.1 Experimental

The procedures for ZIF-8 sensing layer coating and gas test measurements are established and can be found in the literature [67]. Briefly, a gold-coated AT-cut quartz crystal (purchased from INFICON) of resonance frequency = 5 MHz, crystal diameter 25.4 mm, and piezoelectrically active area = 34.19 mm²) was cleaned with piranha solution. Then, ZIF-8 films of 500 nm thickness were coated on all sides of the substrates at room temperature by five repetitive cycles of a dip coating method. The gas testing measurements were performed in a 1000 mL gas cell connected to an automated mass flow controller, as shown in Figure 37.

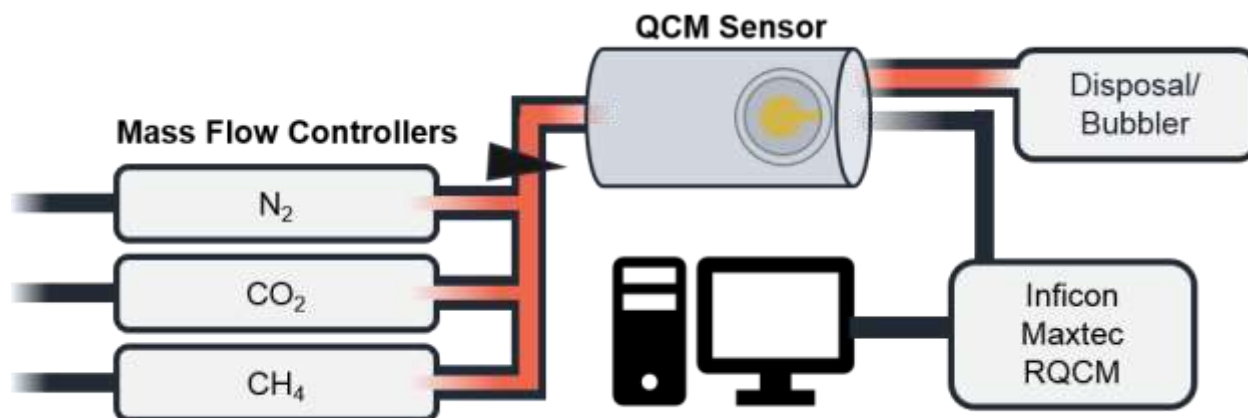


Figure 37. Schematic of experimental setup.

Measurements were taken for gas mixtures of CH_4 , CO_2 , and N_2 , for the component ratios shown in Table 3, below.

Table 3. All gas mixtures tested in this study, in mole fractions.

#	N ₂	CO ₂	CH ₄
1	0.8	0.2	0
2	0.6	0.4	0
3	0.4	0.6	0
4	0.2	0.8	0
5	0	1	0
6	0.8	0	0.2
7	0.6	0.2	0.2
8	0.4	0.4	0.2
9	0.2	0.6	0.2
10	0	0.8	0.2
11	0.6	0	0.4
12	0.4	0.2	0.4
13	0.2	0.4	0.4
14	0	0.6	0.4
15	0.4	0	0.6
16	0.2	0.2	0.6
17	0	0.4	0.6
18	0.2	0	0.8
19	0	0.2	0.8
20	0	0	1

Data was recorded as the shift in frequency due to the exposure of the QCM to each gas mixture. Measurements were taken at 298K and at ambient pressure.

5.2.2 Simulations

As in previous work, GCMC simulations were performed to obtain adsorption of the gas mixtures of N₂, CH₄, and CO₂, of 20% increments, in ZIF-8 at 298K and 1 bar (Table 1). An image of the ZIF-8 unit cell can be seen in Figure 38, below.

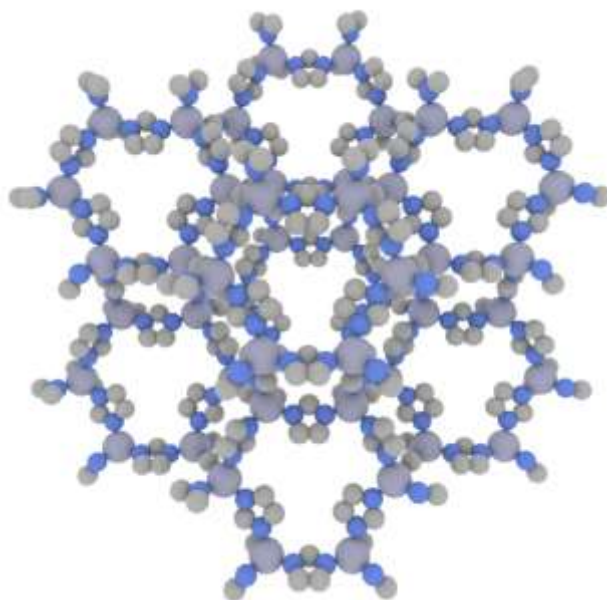


Figure 38. Image of the MOF ZIF-8, chemical structure $\text{Zn}(\text{MeIM})_2$.

The software package RASPA was used, along with the universal force field (UFF) to define structure atoms, as well as the TraPPE force field parameters to define gas molecules [88,89,96]. The validity of this approach, including software and forcefields, has been proven in prior studies [95]. Further simulation details may be found in Appendix A.

As a result, we obtain the total mass adsorbed onto the MOF surface upon exposure to each gas mixture. This mass change is converted to a frequency shift via the Sauerbrey equation (Equation 5-1), used to compare with experimental QCM signals [113].

$$\frac{\Delta m}{A} = \frac{c_m(f-f_0)}{2f^2} \quad (5-1)$$

In Equation 5-1 $\Delta m/A$ (g/cm^2) is the mass change per area, C_m is the mechanical coefficient for quartz ($\text{g}/\text{cm}^2/\text{s}$), f_0 is the fundamental resonance frequency in reference state (5 MHz), and f is the output frequency shift (Hz). Thus, we can use molecular simulations to predict a response for a MOF/QCM sensor.

5.3 Results and Discussion

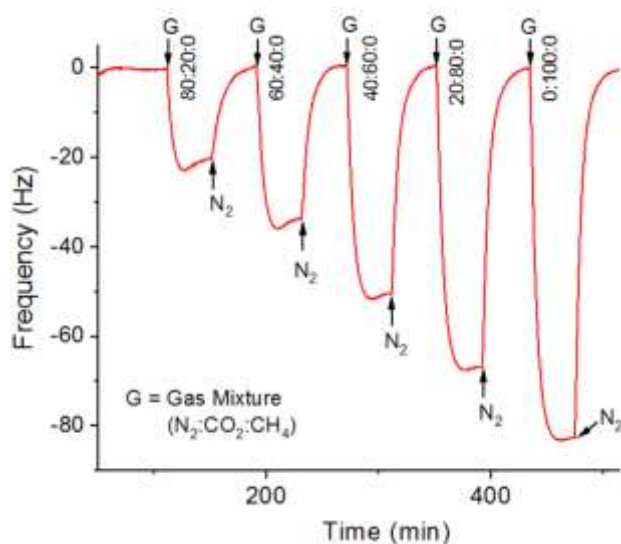


Figure 39. Plot of frequency shift versus time for real time measurements of five gas mixtures of N_2 , CO_2 , and CH_4 .

The real time frequency shift measurements of gases 1-5 (Table 3) are shown in Figure 39, below. Frequency shifts derived from simulations are shown in Figure 40, along with the frequencies from the real time experimental data shown above. The two data sets follow similar trends relative to each gas but are not in agreement with their differing slopes, which is to be expected given the comparison of a defect-free and solvent-free MOF in the computational model

versus the experimental material. Furthermore, the forcefield parameters used in simulations may not perfectly capture interactions with the ZIF-8 framework, however, this computational approach has proven accurate for predicting gas behavior in many MOFs [85,95]. This highlights the need for improved forcefields to develop accurate predictions of MOF sensor behavior.

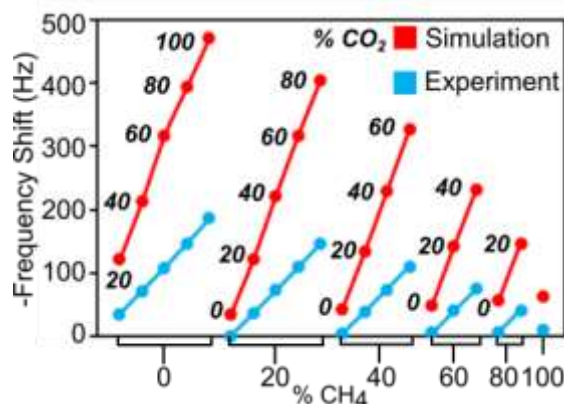


Figure 40. Negative frequency shift vs gas mixture for various concentrations of N₂, CO₂, and CH₄.

ZIF-8 exhibits selectivity towards CO₂, while changes in CH₄ and N₂ are more challenging to detect, as has been shown in the literature [67]. With the interest of natural gas leaks and environmental monitoring in mind, it is a difficult but necessary task to optimize detection of trace amounts of CH₄. In future work, we propose arrays of sensors (i.e. electronic noses) will be effective in improving the overall sensitivity to methane, where we can inform the selection of MOFs using simulations.

5.4 Conclusion

In this work, we have validated the potential for MOFs as sensing materials to integrate with acoustic devices, like QCMs for gas sensing. Furthermore, computational tools are appropriate for predicting mass loading in these types of sensors and therefore may be used to help develop better gas devices. We can expand this method to surface acoustic wave devices [\[25\]](#), and apply this analysis with a variety of MOF materials. However, a limited number of MOFs are known to be able to be deposited on SAWs as thin layers, with parameters like the dielectric constant and elastic modulus not readily available. Such physical properties must be known in order to fabricate SAW devices with MOF films. Future work will focus on MOF arrays that can be applied as electronic noses for the detection of trace gases in complex mixtures.

6.0 The Future of MOF-Based Electronic Nose Development

As discussed in the first section of this work, the methods and analyses presented here are the first of their kind; moreover, they are the first to predict MOF sensor array behavior and systematically investigate array performance. There is much to be done in order to develop optimal electronic noses for today's challenges. I will briefly discuss some future directions based on the work presented here for MOF sensor array development.

6.1 High Throughput Screening of MOFs for Sensing

There are over 6,000 known MOF structures featured in the CoRE MOF database, for which adsorption simulations can be calculated as we have done here. Each one should be considered in the full screening of MOF arrays for sensing. We propose first simulating each structure in RASPA to gather adsorption information for a set of gas mixtures. In continuation of methane-in-air sensing, mixtures of methane, nitrogen, and oxygen may still be considered, but for a large screening less mixtures can be used to filter out poor performing MOFs.

6.1.1 Using physical Parameters as Filtering Criteria

When considering as many as 6,000 MOF structures, using the methods presented here (GCMC simulations, KLD ranking, GA analysis) will not be feasible, considering the amount of possible array combinations to consider. We can think of additional ways to rule out poor

performing MOFs based on their physical properties, like surface area, void fraction, or density. Now that we have predictions for the behavior of 50 MOFs in various array combinations, we can gather information about any trends related to physical properties. For example, if we are looking for the best array to predict methane concentration and all of the known best arrays have a surface area above a certain number, in the larger screening we can rule out any MOFs that have surface areas less than that. A thorough study must be done to draw such conclusions.

6.1.2 Predicting Complex Gas Mixture Adsorption

Currently, as described in this work, the methods we use for predicting gas mixture adsorption in MOFs includes GCMC simulations in the RASPA software package. These simulations require the specifications of energetic parameters for the atoms of both the adsorbent molecules and the porous structures, where established generic force fields have been developed for common molecules (e.g. hydrocarbons, O₂, N₂, etc...). These force field parameters, while adequate for the methods developed here, are inadequate for predictions of complex gas mixtures, especially for disease diagnosis applications. For example, the relevant mixtures for lung cancer diagnosis include over a hundred components; moreover, the molecules of interest cover a range of complex structures that do not have existing force fields. This presents two problems: 1) there are no energetic parameters to simulate these molecules and 2) after three or four components, the GCMC simulations take a long time and become increasingly unreliable, no matter how good the force fields are.

First, for some large, complex molecules there are force fields that work well, so these may be used in simple simulations (i.e. single or binary gas mixtures). Thus, we can do simulations to determine the Henry's constants for these molecules, telling how they adsorb to different MOFs

in the low-pressure region. In the future, the Henry's constants for relevant lung cancer biomarkers will be tested in various concentrations. The challenge remains to accurately predict adsorption of mixtures having a hundred molecules, where Henry's constants can only be used for some. Additionally, more work must be done to develop force fields for these biomarkers, because this challenge is relevant for many disease and biosensing related applications. Until accurate force fields are developed, it will not be possible to predict the interactions of complex mixtures.

6.1.3 Interpolation to Predict Gas Mixture Adsorption

Our current methods for predicting gas mixture adsorption require performing GCMC simulations in the RASPA software package. The simulations are short relative to energetic calculations but require a significant amount of computational time that it is infeasible to simulate every possible gas mixture. One adsorption simulation (1 mixture + 1 MOF) may take anywhere from ten minutes to twenty-four hours, and the simulations only get more complicated as more gases are added to the mixture. Thus, there must be a more efficient way to obtain the adsorption information for complex gas mixtures.

If we have a library of simulation results for ternary gas mixtures in MOFs, it may be possible to use this information to predict adsorption of gas mixtures that were not simulated. We hypothesize that adsorption is mostly linear in small increments and that interpolation may be used to calculate adsorption of a new gas mixture. If a range of mixtures were simulated, then we could obtain adsorption results from gases that are spatially between those mixtures. This would allow us to cover smaller increments of compositions of gas mixtures without the computational burden.

6.1.4 Tiered Approach

The vision for this work is to bring together multiple ranking metrics and approaches, implementing them in a series. In an ideal scenario, we would perform atomistic simulations for every possible MOF-gas mixture combination and look at every array combination to find the best for any application. Unfortunately, this would take more computational time than is feasible, so we have to narrow down the amount of materials we consider when we reach the step of simulating adsorption. Therefore, we can combine some of the proposed methods in this section to intelligently filter and rank MOFs for gas sensing.

For example, given an application of methane leak detection, the first step could eliminate any MOFs that have a surface area outside of a range previously discovered to be the best for methane sensing. This could take you from 6,000 possible MOF structures to 3,000 structures. Then, for those 3,000 structures you can simulate the Henry's constants for methane (a very short simulation) and eliminate any MOFs below a certain threshold. Finally, using the MOFs that are remaining you may perform GCMC simulations using a range of gas mixtures, interpolate between them to gather more adsorption information, and then look at trends in adsorption. It is possible to also rank MOFs for their sensitivity and selectivity to methane in the presence of other gases. The use of the SAGS score may then be appropriate to look at MOF array sensitivity to a range of gas mixtures. Once more MOFs have been eliminated, finally they can be tested in arrays according to the methods described here, including the GA analysis and KLD ranking metric.

6.2 Alternative Signal Transduction Mechanisms

6.2.1 Electrical Conductivity

There exist a handful of MOFs that have been known to show properties of electrical conductivity, which is an attractive type of signal transduction mechanism because the responses tend to be highly sensitive.[\[63\]](#) Just as with conducting polymers, the existence of conductive MOFs is a relatively rare occurrence, and little is currently known about their mechanisms of operation. This lends itself toward much opportunity for gas sensor improvements using these materials, where electronic nose signals could be greatly amplified, but much work must be done until practical implementation is possible.

Along with predicting adsorption of gases in MOF structures, it is possible to calculate conductivity of materials using density functional theory (DFT). DFT is widely used by researchers to perform energetic and structural optimization calculations for various applications. Such simulations are non-trivial and are computationally intensive, requiring significant development of the computational approach to ensure accuracy. Currently, there are experimental conductivity measurements available in published work, so the DFT calculations are being optimized by doctoral students in the Wilmer lab. Once this method is established, the prediction of conductivity in any MOF will be possible, so this will be an additional tool for screening material properties.

In the implementation of conductive MOFs in electronic noses will provide the benefit of not only a sensitive transduction mechanism but will allow for multiple transduction mechanisms in a single device. The optimal electronic nose would employ as many signal types as possible, maximizing signal diversity and therefore providing the most information.

6.2.2 Optical

An additional transduction mechanism that can be used with MOFs is an optical response, or a change in the material's color. This can occur through a process called solvatochromism, which is a shift in the absorption spectrum of a material in response to a change in the surrounding solvent.[56] For example, it has been shown that some MOFs containing Co^{2+} metal nodes have a color change response upon exposure chlorine containing compounds.[56]

MOFs also exhibit another optical response in the form of photoluminescence, occurring during quenching or enhancing of photoinduced emission as a result of guest adsorption.[56] Luminescent MOFs have been one of the most widely studied for sensing, which may lend itself well to predicting luminescent behavior in new MOF structures. In future work, all these types of signal transduction should be used in conjunction with each other (e.g. optical, conductive, mass-based). Machine learning techniques may be used to predict signals of various types upon exposure to different gases.

6.3 Electronic Nose Architecture

Beyond signal transduction mechanisms, another way to advance electronic nose technology is through innovation of the device itself. Many strides have been made to optimize the sensitivity of device platforms, and work is continuing to be done to improve compatibility of new materials with devices. We can propose here some electronic nose configurations that correspond with the work presented regarding MOF array sensing.

First, MOFs, in correlation with their sensing capabilities, have known high gas storage and separation capabilities, where they are being studied for membranes and environmental applications. Additionally, a common challenge with complex sensing is the presence of interferences that inhibit sensitivity and selectivity towards target gases. For example, methane sensing remains a challenge due to carbon dioxide and water interference. One can then imagine a scenario where such gases could be filtered out via separation by a MOF targeted for a specific interferent (e.g. carbon dioxide). Next, the remaining gas could be filtered further or sent through to the sensing platform(s). Some current devices include pumps and fans, but these may be bulky and inadequate for portable sensing. In this scenario, we would aim to build a portable device that can perform such actions with low power consumption.

Further, we can propose various configurations of layering sensing materials. If we have information from this computational optimization process to inform MOF selection in arrays, then we could potentially select specific MOFs with complimentary properties. With advances in sensing layer deposition technology, we could layer different materials in an intelligent manner, where the combination of their sensing capabilities works to amplify the sensitivity and selectivity of the sensor. From a simulation perspective, we could use the same atomistic adsorption calculations we currently perform, it would require development of data analysis techniques post-simulation. Additionally, in this same vein, work has been done to predict the interpenetration of MOFs and such structures may possess amplified gas detection properties. Along with the predictive optimization analysis presented here, the tuning and functionalization of MOFs will lead to advances in sensing technologies.

6.4 Post-Simulation Data Analysis Techniques

As the first steps in the optimization of MOF arrays for gas sensing, this work has limited machine learning techniques to just a genetic algorithm; however, there are many available methods for predictive analyses that can be applied to gas sensing. Many researchers use methods like principal component analysis and neural networks to process sensor signals and create libraries of data, but in our approach, we used similar methods to predict materials' performance before any experimental development. Therefore, next steps in this work would include the use of sophisticated machine learning techniques for predicting MOF array performance.

Although the genetic algorithm works well for the set of fifty MOFs that we tested, as we have mentioned there are upwards of 6,000 available structures. In addition to the previously described tiered approach to selecting an optimal MOF array, we propose predictive methods to further strengthen this methodology. Specifically, the use of artificial neural networks (ANNs) may prove integral to electronic nose optimization. For example, rather than having to identify ranges of physical parameters that benefit an array, these will be resolved without having to explicitly tell the ANN. Using the array data we currently possess, we could input any number of identifiers for MOFs, including atom positions, surface area, or void fraction. Then the neural network would be trained to identify arrays with high KLD values, or even arrays which can resolve a particular gas with great sensitivity and selectivity.

The future of MOF based electronic nose development will require layers of advanced simulation and data analysis techniques. If this research follows in its current trajectory, in the next decade we may see the most advanced gas sensing technology covering a vast amount of critical applications.

Appendix A Computing Gas Mixture Adsorption

Adsorption data was obtained using the RASPA software package, developed by Dubbledam et al.[96] Grand canonical Monte Carlo (GCMC) simulations were performed with 1,326 different gas mixtures of methane, nitrogen, and oxygen, varying their concentrations from zero to one by 2%. We simulated each mixture with fifty MOF structures; nine of which were used in a previous study[108] and forty-one of which were selected from the CoRE MOF database developed by Chung et al.[107] The surface area (SA), density, and void fraction of each MOF are listed below in Table 4, as well as the unit cell parameters specified in each input file. Selected MOFs have their common names also listed in parentheses.

Table 4. Physical properties of MOF structures

MOF	SA (m ² /g)	Density (g/cm ³)	Void Fraction	Unit Cells (a b c)
IRMOF-1	25173	0.5934	0.825	1 x 1 x 1
HKUST-1	1500-2100	0.8791 ^{6,7}	0.765	1 x 1 x 1
NU125	31208	0.5784	0.859	1 x 1 x 1
ZIF-8	181310	0.9246	0.212	2 x 2 x 2
UIO-66	118713	1.2191	0.526	2 x 2 x 2
Mg-MOF-74	1200-1525	0.9111	0.565	2 x 2 x 4
NU-100	614322	0.2912	0.821	1 x 1 x 1
MOF-801	69024	1.6824	0.454	2 x 2 x 2
MOF-177	3275-4630	1.0113	0.679	1 x 1 x 1
ALUKIC	5132	0.567	0.83	2 x 2 x 1
AMIMAL	1141	0.989	0.56	2 x 1 x 1
AXUHEH	825	1.065	0.35	2 x 2 x 1
BAZGAM	6581	0.127	0.97	1 x 1 x 1
BIWSEG	2267	0.467	0.81	1 x 1 x 1
EDUVOO	4857	0.373	0.91	2 x 2 x 2
FIDRIV	1893	0.698	0.69	2 x 2 x 2

Table 4 (continued)

GAGZEV	5777	0.279	0.92	1 x 1 x 1
GUPBEZ01	651	2.454	0.20	2 x 2 x 2
HABQUY (PCN-610)	5750	0.289	0.91	1 x 1 x 1
HIFTOG	1885	1.166	0.57	2 x 2 x 2
JEWCAP	623	1.115	0.27	2 x 2 x 1
KICXAX	72	3.585	0.15	2 x 2 x 2
KIFJUF	2654	0.822	0.59	3 x 2 x 2
KINKAV	190	1.210	0.19	3 x 2 x 2
LODPUQ	1210	1.074	0.52	2 x 2 x 2
LOFVUY	1559	1.078	0.59	2 x 1 x 1
MUDTEL	3719	0.559	0.85	1 x 1 x 1
NAYZOE (Cu-TCA)	4640	0.499	0.86	2 x 2 x 2
NIBHOW	5103	0.280	0.92	1 x 1 x 1
NIBJAK	5417	0.223	0.94	1 x 1 x 1
OFEREX	904	1.568	0.61	3 x 3 x 2
RAVXET (IRMOF-74-VI-oeg)	2809	0.327	0.79	4 x 1 x 1
RAVXIX (IRMOF-74-IX)	3036	0.235	0.86	4 x 1 x 1
RAVXOD (IRMOF-74-XI)	3299	0.179	0.88	4 x 1 x 1
RUTNOK (IRMOF-76)	6200	0.241	0.90	1 x 1 x 1
SADLEQ	930	1.505	0.60	3 x 3 x 2
SAPBIW	2994	0.306	0.89	1 x 1 x 1
SICZOV	4151	0.420	0.90	2 x 2 x 2
TOHSAL (meso-[LCu ₂ (H ₂ O) ₂])	4312	0.576	0.80	2 x 2 x 1
UKUPUL	881	1.434	0.56	2 x 2 x 2
VETTIZ	2060	0.538	0.83	1 x 1 x 1
WIYMOG	6833	0.408	0.81	2 x 2 x 1
WUNSEE01	411	1.209	0.39	2 x 2 x 2
XAFFAN	5181	0.365	0.89	2 x 2 x 2
XAFXOT	280	1.888	0.24	6 x 3 x 2
XAHQAA	6250	0.170	0.95	1 x 1 x 1
XALTIP	3296	0.551	0.85	2 x 2 x 2
XUKYEI	6327	0.287	0.88	2 x 2 x 2
XUWVUG	20	3.194	0.10	7 x 2 x 2
YEQRIV	3528	0.742	0.65	3 x 2 x 2

Simulations were carried out at a pressure of 1 bar, 298K and using 2,000 production cycles with 1,000 initialization cycles. A cycle consists of n Monte Carlo steps; where n is equal to the number of molecules in the simulation domain (which fluctuates during a GCMC simulation). Simulations also included random insertion, deletion, and translation of molecules with equal probability. Rigid MOF structures were assumed, and a Lennard-Jones (LJ) cutoff of 12 Å was used, with the number of unit cells used for each MOF structure as specified in Table 4, for each simulation. The LJ potential is calculated during simulations to determine the overall energy of the structure with adsorbed gases,

$$V_{ij} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (\text{A-1})$$

requiring the use of the parameters of well-depth, ϵ , and radius, σ , as seen in the above LJ equation (Equation A-1).

Table 5. LJ Parameters for Gas Molecules

Atom Type	ϵ/k_B [K]	σ [Å]
CH ₄	148	3.73
O_CO ₂	79	3.05
C_CO ₂	27	2.8
O_O ₂	49	3.02
N_N ₂	36	3.31
N COM	0	0
CH ₃ _sp ³	98	3.75

The LJ parameters for atoms of the gas mixtures were obtained from construction of the TraPPE force field by Martin et al.[88] for methane and ethane, Potoff et al.[90] for CO₂ and N₂, and Zhang et al.[114] for O₂. The radius and well-depth from the TraPPE force field for the gases used are shown in Table 5. LJ parameters for MOF atoms were obtained from the Universal Force Field developed by Rappe et al.[89], for both metal atom types and organic atoms, as shown in Table 6.

Table 6. LJ parameters for framework atoms in all MOFs

Atom Type	ϵ/k_B [K]	σ [Å]
C	52.84	3.851
O	30.19	3.5
H	22.14	2.886
N	34.72	3.66
Zr	34.72	2.783
Zn	62.4	2.462
Cu	2.52	3.114
Mg	55.86	2.691
As	155.47	3.77
Ce	6.54	3.17
Dy	3.52	3.05
Eu	4.03	3.11
I	170.57	4.01
K	17.61	3.4
La	8.55	3.14
Na	15.09	2.66
Nd	5.03	3.18
Tb	3.52	3.07
W	33.71	2.73

The partial charges of the gas molecules used by RASPA to carry out GCMC simulations are listed in Table 7, N₂ obtained from Potoff et al.[[90](#)] and O₂ from Zhang et al.[[114](#)], as used for the LJ parameters obtained from the TraPPE force field.

Table 7. Partial charges for gas molecule atoms

Atom Type	q(e)
CH ₄	0
O_CO ₂	-0.35
C_CO ₂	0.7
O_O ₂	-0.113
O COM	0.226
N_N ₂	-0.482
N COM	0.964
CH ₃ _sp3	0

Table 8. Partial charges of framework atoms

Average Partial Charges (1.60218 e-19 C/particle)									
Atom Type	IRMOF-1	HKUST-1	NU-125	UiO-66	ZIF-8	MgMOF-74	NU-100	MOF-177	MOF-801
H	0.054	0.044	0.069	-	0.128	0.0581	0.0647	0.035	-0.0551
C	0.053	0.107	0.043	0.123	-0.142	0.0831	0.0108	0.0179	-0.258
N	-	-	-0.074	-	-0.318	-	-	-	-
O	-0.52	-0.4	-0.385	-0.782	-	-0.621	-0.332	-0.528	-0.175
Cu	-	0.867	0.9	-	-	-	0.927	-	-
Zn	1.212	-	-	-	1.13	-	-	1.211	-
Zr	-	-	-	3.187	-	-	-	-	2.541
Mg	-	-	-	-	-	1.471	-	-	-

The partial charges of the MOF framework atoms were obtained using the EQueq charge equilibration method developed by Wilmer et al.[115] The average charge per atom for the first nine MOFs are listed in Table 8. The full atomic descriptions (cif files) may be found in the Supporting Information Files of published manuscripts. The methane molecules were modeled as single spheres, so their atom positions are not applicable, just as their charges were zero overall for a single sphere. The remaining gas molecules, oxygen [114] and nitrogen,[116] are both modeled as rigid three atom structures with their positions listed in Table 9.

Table 9. Atom positions for 3 dimensional molecules

	X	Y	Z
O₂			
0 O_o2	0	0	0.605
1 O_com	0	0	0
2 O_o2	0	0	-0.605
N₂			
0 N_n2	0	0	0.55
1 N_com	0	0	0
2 N_n2	0	0	-0.55
CO₂			
0 O_co2	0	0	1.16
1 C_co2	0	0	0
2 O_co2	0	0	-1.16

The Peng-Robinson equation of state was used to calculate fugacities needed to run the GCMC simulations, shown in Equations A-2 through A-7, below.

$$p = \frac{RT}{V_m - b} - \frac{a\alpha}{V_m^2 + 2bV_m - b^2} \quad (\text{A-2})$$

$$a = \frac{0.457235 R^2 T_c^2}{p_c} \quad (\text{A-3})$$

$$b = \frac{0.077796 R T_c}{p_c} \quad (\text{A-4})$$

$$\alpha = \left(1 + \kappa(1 - T_r^{0.5})\right)^2 \quad (\text{A-5})$$

$$\kappa = 0.37464 + 1.54226\omega - 0.26992\omega^2 \quad (\text{A-6})$$

$$T_r = \frac{T}{T_c} \quad (\text{A-7})$$

The critical parameters for the simulated gas molecules, methane, oxygen, and nitrogen are listed in Table 10.

Table 10. Critical constants for gas molecules, as seen in the PR EoS

Molecule	T _c [K]	P _c [MPa]	ω
CH ₄	190.56	4.60	0.011
CO ₂	304.13	7.38	0.224
O ₂	154.6	5.05	0.022
N ₂	126.19	3.40	0.037
C ₂ H ₆	305.33	4.87	0.099

Appendix B Lists of Simulated Gas Mixtures

B.1 Sensor Array Gas Space Score

Table 11. List of all gas mixtures simulated for SAGS score study, in mole fractions

	N ₂	CH ₄	CO ₂	C ₂ H ₆		N ₂	CH ₄	CO ₂	C ₂ H ₆		N ₂	CH ₄	CO ₂	C ₂ H ₆
1	0.4	0.575	0	0.025	27	0.5	0.12	0.3	0.08	53	0.65	0	0.25	0.1
2	0.4	0.45	0.1	0.05	28	0.5	0.07	0.34	0.09	54	0.7	0.28	0	0.02
3	0.4	0.325	0.2	0.075	29	0.5	0.02	0.38	0.1	55	0.7	0.21	0.05	0.04
4	0.4	0.2	0.3	0.1	30	0.5	0	0.38	0.12	56	0.7	0.16	0.09	0.05
5	0.4	0.175	0.3	0.125	31	0.55	0.44	0	0.01	57	0.7	0.11	0.13	0.06
6	0.4	0.1	0.35	0.15	32	0.55	0.38	0.05	0.02	58	0.7	0.06	0.17	0.07
7	0.4	0	0.425	0.175	33	0.55	0.33	0.09	0.03	59	0.7	0.02	0.2	0.08
8	0.45	0.54	0	0.01	34	0.55	0.28	0.13	0.04	60	0.7	0	0.2	0.1
9	0.45	0.45	0.08	0.02	35	0.55	0.21	0.19	0.05	61	0.75	0.23	0	0.02
10	0.45	0.44	0.08	0.03	36	0.55	0.16	0.23	0.06	62	0.75	0.16	0.05	0.04
11	0.45	0.415	0.095	0.04	37	0.55	0.12	0.26	0.07	63	0.75	0.11	0.08	0.06
12	0.45	0.4	0.1	0.05	38	0.55	0.1	0.27	0.08	64	0.75	0.07	0.11	0.07
13	0.45	0.32	0.17	0.06	39	0.55	0.05	0.31	0.09	65	0.75	0.02	0.15	0.08
14	0.45	0.28	0.2	0.07	40	0.55	0	0.35	0.1	66	0.75	0	0.15	0.1
15	0.45	0.2	0.27	0.08	41	0.6	0.38	0	0.02	67	0.8	0.18	0	0.02
16	0.45	0.14	0.32	0.09	42	0.6	0.26	0.1	0.04	68	0.8	0.11	0.05	0.04
17	0.45	0.1	0.35	0.1	43	0.6	0.17	0.17	0.06	69	0.8	0.05	0.09	0.06
18	0.45	0.05	0.4	0.1	44	0.6	0.07	0.25	0.08	70	0.8	0.02	0.1	0.08
19	0.45	0	0.43	0.12	45	0.6	0	0.3	0.1	71	0.8	0	0.1	0.1
20	0.5	0.49	0	0.01	46	0.65	0.33	0	0.02	72	0.85	0.11	0	0.04
21	0.5	0.43	0.05	0.02	47	0.65	0.26	0.05	0.04	73	0.85	0.05	0.04	0.06
22	0.5	0.37	0.1	0.03	48	0.65	0.2	0.1	0.05	74	0.85	0.02	0.05	0.08
23	0.5	0.31	0.15	0.04	49	0.65	0.14	0.15	0.06	75	0.85	0	0.05	0.1
24	0.5	0.27	0.18	0.05	50	0.65	0.1	0.18	0.07	76	0.9	0.04	0	0.06
25	0.5	0.22	0.22	0.06	51	0.65	0.06	0.21	0.08	77	0.9	0	0.02	0.08
26	0.5	0.18	0.25	0.07	52	0.65	0.03	0.23	0.09	78	0.9	0	0	0.1

Calculations for the *sensor array gas space* (SAGS) score were calculated using a Mathematica code and the adsorption results obtained from GCMC simulations in Raspa software package. The SAGS score requires the composition values and total mass adsorbed for each MOF-gas mixture combination. As mentioned in the main text, our study consisted of 78 different gas mixtures (seen in Table 11) and 5 different MOFs (IRMOF-1, HKUST-1, NU-125, UiO-66, ZIF-8), resulting in a total of 390 simulation results.

B.2 Kullback-Liebler Divergence Calculation

Table 12. List of all gas mixtures simulated for information content optimization study, in mole fractions

	CH₄	N₂	O₂
1	0.0	0.01	0.99
2	0.0	0.02	0.98
3	0.0	0.03	0.97
4	0.0	0.04	0.96
5	0.0	0.05	0.95
6	0.0	0.06	0.94
7	0.0	0.07	0.93
8	0.0	0.08	0.92
9	0.0	0.09	0.91
10	0.0	0.10	0.90
⋮	⋮	⋮	⋮
5,151	1.0	0.0	0.0

For the demonstration of the Kullback-Liebler divergence calculation, ranking arrays of MOFs for their ability to resolve mixtures of CH₄, N₂, and O₂. The simulated mixtures covered a range of compositions from 0 to 1 for each gas, varying mole fractions by 1% increments. The result was 5,151 mixtures, as demonstrated in Table 12.

B.3 Genetic Algorithm Analysis

In a similar manner to the previous section, the gas mixtures for the genetic algorithm analysis were taken for gases CH₄, N₂, and O₂, varying their concentrations from 0 to 1 by 2% increments.

Table 13. List of all gas mixtures simulated for SAGS score study, in mole fractions

	CH₄	N₂	O₂
1	0.0	0.02	0.98
2	0.0	0.04	0.96
3	0.0	0.06	0.94
4	0.0	0.08	0.92
5	0.0	0.10	0.90
6	0.0	0.12	0.88
7	0.0	0.14	0.86
8	0.0	0.16	0.84
9	0.0	0.18	0.82
10	0.0	0.20	0.80
⋮	⋮	⋮	⋮
1,326	1.0	0.0	0.0

Appendix C Sensor Array Gas Space Score

C.1 Example Calculation

The calculation of the SAGS score is carried out below, for a gas space of just three gas mixtures (as opposed to 78 in the full study) and a sensor array of 3 MOFs: IRMOF-1, HKUST-1, NU-125. We will take the first three gas mixtures from our study (Table 14):

Table 14. Three gas mixture compositions and total masses adsorbed in three different MOFs

Mixture #	N ₂	CH ₄	CO ₂	C ₂ H ₆	IRMOF-1 (mg/cc framework)	HKUST-1 (mg/cc framework)	NU-125 (mg/cc framework)
1	0.4	0.575	0	0.025	6.653	35.61	19.59
2	0.4	0.45	0.1	0.05	12.25	57.58	35.77
3	0.4	0.325	0.2	0.075	17.74	81.18	51.45

First, we take the pairwise array (PA) scores for each combination of pairs of gas mixtures that we have simulated. For this example, we take three PA scores, for mixtures 1 and 2, 1 and 3, and 2 and 3. Equation C-1 is used to calculate the compositional distances between each pair of gas mixtures, i and j .

$$d_{ij} = \sqrt{\sum_{k=1}^N (x_{k,i} - x_{k,j})^2} \quad (\text{C-1})$$

$$d_{12} = \sqrt{(x_{N2,1} - x_{N2,2})^2 + (x_{CH4,1} - x_{CH4,2})^2 + (x_{CO2k,1} - x_{CO2,2})^2 + (x_{C2H6,1} - x_{C2H6,2})^2}$$

$$d_{13} = \sqrt{(x_{N2,1} - x_{N2,3})^2 + (x_{CH4,1} - x_{CH4,3})^2 + (x_{CO2k,1} - x_{CO2,3})^2 + (x_{C2H6,1} - x_{C2H6,3})^2}$$

$$d_{23} = \sqrt{(x_{N2,2} - x_{N2,3})^2 + (x_{CH4,2} - x_{CH4,3})^2 + (x_{CO2k,2} - x_{CO2,3})^2 + (x_{C2H6,2} - x_{C2H6,3})^2}$$

$$d_{12} = \sqrt{(0.4 - 0.4)^2 + (0.575 - 0.45)^2 + (0.0 - 0.1)^2 + (0.025 - 0.05)^2}$$

$$d_{13} = \sqrt{(0.4 - 0.4)^2 + (0.575 - 0.325)^2 + (0.0 - 0.2)^2 + (0.025 - 0.075)^2}$$

$$d_{23} = \sqrt{(0.4 - 0.4)^2 + (0.45 - 0.325)^2 + (0.1 - 0.2)^2 + (0.05 - 0.075)^2}$$

$$d_{12} = 0.1620$$

$$d_{13} = 0.3240$$

$$d_{23} = 0.1620$$

Next, we calculate the mass difference for each pair of gas mixtures, using Equation C-2, where k is one MOF, for M MOFs in the sensor array.

$$m_{ij} = \sqrt{\sum_{k=1}^M (m_{k,i} - m_{k,j})^2} \quad (\text{C-2})$$

$$m_{12} = \sqrt{(m_{IRMOF-1,1} - m_{IRMOF-1,2})^2 + (m_{HKUST-1,1} - m_{HKUST-1,2})^2 + (m_{NU-125,1} - m_{NU-125,2})^2}$$

$$m_{13} = \sqrt{(m_{IRMOF-1,1} - m_{IRMOF-1,3})^2 + (m_{HKUST-1,1} - m_{HKUST-1,3})^2 + (m_{NU-125,1} - m_{NU-125,3})^2}$$

$$m_{23} = \sqrt{(m_{IRMOF-1,2} - m_{IRMOF-1,3})^2 + (m_{HKUST-1,2} - m_{HKUST-1,3})^2 + (m_{NU-125,2} - m_{NU-125,3})^2}$$

$$m_{12} = \sqrt{(6.653 - 12.25)^2 + (35.61 - 57.58)^2 + (19.59 - 35.77)^2}$$

$$m_{13} = \sqrt{(6.653 - 17.74)^2 + (35.61 - 81.18)^2 + (19.59 - 51.45)^2}$$

$$m_{23} = \sqrt{(12.25 - 17.74)^2 + (57.58 - 81.18)^2 + (35.77 - 51.45)^2}$$

$$m_{12} = 27.85$$

$$m_{13} = 56.70$$

$$m_{23} = 28.86$$

Then, we calculate the PA scores for each gas mixture combination using Equation C-3:

$$S_{ij} = \frac{m_{ij}}{d_{ij}} \quad (C-3)$$

$$S_{12} = \frac{m_{12}}{d_{12}} \quad S_{13} = \frac{m_{13}}{d_{13}} \quad S_{23} = \frac{m_{23}}{d_{23}}$$

$$S_{12} = \frac{27.85}{0.162} \quad S_{13} = \frac{56.70}{0.324} \quad S_{23} = \frac{28.86}{0.162}$$

$$S_{12} = 171.9 \quad S_{13} = 175.0 \quad S_{23} = 178.2$$

Lastly, we take the overall SAGS score by determining the average of all of the PA scores, for W combinations of pairs of gas mixtures (Equation C-4):

$$\phi_W = \frac{\sum S_{ij}}{W} \quad (C-4)$$

$$\phi_3 = \frac{S_{12} + S_{13} + S_{23}}{3} = \frac{171.9 + 175.0 + 178.2}{3} = 175.0$$

One can see that the SAGS score for a sensor array consisting of IRMOF-1, HKUST-1, and NU-125 yields a score of 175.0 for distinguishing the three gas mixtures chosen in the analysis. This same sensor array yields a score of 0.1283 at 1 bar for distinguishing all 78 of the gas mixtures used in our full analysis.

C.2 List of All Results

Below is a list of all sensor arrays for which SAGS scores were calculated (Table 15). All combinations of MOFs were taken, ranging from 1 to 5 MOFs in an array. Mixture numbers correspond to those found in Figure 5 (main text).

Table 15. List of All MOF combinations and their scores, ranging from 1-5 MOF arrays. Key: IRMOF-1 (I), HKUST-1 (H), NU-125 (N), UiO-66 (U), ZIF-8 (Z)

Mixture #	1 MOF	SAGS score	
		1 bar	10 bar
1	IRMOF-1	0.0255269	0.275097
2	NU-125	0.0698682	0.269723
3	HKUST-1	0.103688	0.284623
4	UiO-66	0.106894	0.29162
5	ZIF-8	0.113112	0.142338
Mixture #	2 MOF	1 bar	10 bar
6	IN	0.0746132	0.389617
7	IH	0.106924	0.403791
8	IU	0.110437	0.414541
9	IZ	0.116442	0.315399
10	HN	0.125658	0.394177
11	NU	0.13011	0.405484
12	NZ	0.133939	0.307165
13	HU	0.151354	0.412617
14	HZ	0.155753	0.319187
15	UZ	0.159746	0.327707

Mixture #	3 MOF	SAGS score	
		1 bar	10 bar
16	IHN	0.12831	0.486939
17	INU	0.132757	0.49815
18	INZ	0.136612	0.417617
19	IHU	0.153606	0.506432
20	IHZ	0.157977	0.429875
21	IUZ	0.161964	0.440401
22	HNU	0.167632	0.497094
23	HNZ	0.171065	0.420122
24	NUZ	0.175172	0.430984
25	HUZ	0.191953	0.437588
Mixture #	4 MOF	1 bar	10 bar
26	IHNU	0.169643	0.575996
27	IHNZ	0.173074	0.508739
28	INUZ	0.177154	0.51949
29	IHUZ	0.193741	0.527236
30	HNUZ	0.20474	0.517921
Mixture #	5 MOF	1 bar	10 bar
31	IHNUZ	0.20474	0.594306

Appendix D Kullback-Liebler Divergence

D.1 Detailed Explanation of Algorithm

First, we began with a set of materials, nine different MOF structures, as well as gases of interest, CH₄, N₂, and O₂. We decided that to cover a large range of gases that we wanted all gas mixtures containing the three gases varying by 1%, from 0 to 1 mole fraction of each. This resulted in 5,151 total gas mixtures, to be simulated in nine MOFs. GCMC simulations were performed using the RASPA software package, resulting in total adsorption (mg/cm³) of each gas mixture in each MOF, at 298K and 1 bar (5,151 x 9 = 46,359 simulations).

The simulation output files from RASPA were processed before being input into our main algorithm. Since we first analyzed array performance for binary mixtures of CH₄ and N₂, we selected only the mixtures where O₂ had a mole fraction of 0 and used those in the analysis. Then, we included all the mixtures in the ternary analysis, but the process overall is the same regardless of the number of components in the mixture. Files were set up as inputs to the algorithm for each “experiment” tested (see description of experimental/unknown gas mixtures in main text), one with the experimental mixtures and another with the simulation results.

Each experiment requires separate input files, with the experimental file containing the mixture and total adsorption for each MOF (9 rows). The simulation results file lists each MOF/mixture/adsorption from the simulations, after the entries being used as the experiment are taken out (46,359 – 9 = 46,350 rows). For the ternary mixtures, we tested 9 “experiments”, and for each experiment we had two input files, resulting in a total of 18 different input files for the ternary analysis.

Each experiment was then tested individually through the following algorithm. For each MOF, the “experimental” mass value is taken as the mean of a normal distribution. The distribution must be truncated so that no negative mass values are possible, so the x-axis begins at 0 (0 total mass adsorption). Given the parameters of mean total adsorption and a standard deviation of 10%, we are testing how close all the simulated masses are to the “experimental” mass. For each gas mixture in the current MOF, the probability for the corresponding simulated adsorption (5,150 per MOF) value is calculated per this normal distribution. Essentially, all the mixtures with adsorption closest to the experimental adsorption will have higher probabilities, and vice versa. The calculated probabilities are in the form of probability mass function (pmf), so after all probabilities for a MOF are determined, they are normalized so that the sum over all pmf values is 1. This process results in a pmf value corresponding to each gas mixture, calculated separately for all MOFs.

Next, we have all probabilities for all MOFs so it is time to calculate probabilities for arrays of MOFs. All combinations are taken, from 1-9 MOFs in an array, and joint probabilities are determined for each array/gas. The joint probability is simply taking the pmf values for all MOFs in that array and multiplying them pointwise and normalizing the result to the sum of pmfs over all concentrations is 1. Now, we have pmf vs concentration for each MOF array combination. From here we can quantify and compare their performances for determining the “experimental” mixture.

Considering the ternary mixture case, we now analyze the behavior of MOFs as they relate to each component in the mixture. So, for each gas/MOF combination there is a set of mole fractions and corresponding pmf values. If we were to plot all 5,150 pmf vs concentration values for each gas, the curve would not be clear, since each mole fraction will have multiple pmfs from different mixtures. For example, a mixture of CH₄: 0.2, N₂: 0.3, O₂: 0.5 and CH₄: 0.2, N₂: 0.4, O₂: 0.4 will have different pmfs but when isolating CH₄, those values would be plotted at the same

mole fraction. This problem is mitigated when we combine pmf values at each mole fraction for each gas, by taking the sum. This results in 101 mole fractions (0-1) for each gas, having a pmf value for each MOF array associated with that concentration.

Finally, for each array/gas combination we calculate the Kullback-Liebler Divergence (KLD) value, in terms of bits of information. This information content is a relation between the pmfs (P_i) we calculated and a reference probability ($Q_i = 1/101$). This KLD value is calculated per Equation 6 in the main text, resulting in a KLD for each array/gas combination. The arrays are then ranked, the best having the highest KLD and the worst with the lowest. The main difference in analysis between the binary and ternary mixtures is the variations in KLD. For binary mixtures, the KLD for each gas is the same, because if an array can predict one concentration with some certainty, then by default it must be just as certain about the other component's concentration. On the other hand, for ternary mixtures the best/worst MOF arrays are not as obvious because of varying KLDs between gases. So, we multiply the KLD values from each gas for an array and use those values to rank the best/worst arrays in the ternary analysis.

The probability plots in the manuscript feature the probability density vs concentration, rather than the pmfs. This is a simple conversion, multiplying each pmf value by the total number of points in the distribution (101 for each gas in our case). In the case of the triangle plots from Figure 18 and Figure 19 in the main text, each point was plotted individually without averaging any pmfs, so conversion to probability density was a multiplication of 5,150.

D.2 Example Calculation

First, we have adsorption simulation results for gas mixtures, considering two different MOFs. Table 16 below contains examples of mixtures and their results.

Table 16. Sample list of gas mixtures and simulation results

Component Mole Fraction			Total Mass Adsorbed (mg/cm ³)	
CH ₄	O ₂	N ₂	IRMOF-1	ZIF-8
0.05	0.15	0.8	4.72	10.77
⋮	⋮	⋮	⋮	⋮
0.1	0.5	0.4	5.24	12.60
⋮	⋮	⋮	⋮	⋮
0.3	0.6	0.1	5.47	15.24
⋮	⋮	⋮	⋮	⋮
0.5	0.3	0.2	5.41	16.87
⋮	⋮	⋮	⋮	⋮
0.8	0.1	0.1	5.50	20.23

Next, we select an “experimental” gas mixture which we will determine if we can predict using MOF arrays. In this example, we select Experiment 1 from the text as our unknown gas mixture, and the details for the gas composition and adsorption results are below in Table 17.

Table 17. Experiment 1 composition and simulation results

Component Mole Fraction			Total Mass Adsorbed (mg/cm ³)	
CH ₄	O ₂	N ₂	IRMOF-1	ZIF-8
0.1	0.15	0.75	4.52	11.67

This mixture is taken out of the simulation results and the masses for each MOF are saved as experimental masses. The next step in the calculation is to create truncated normal distributions, using each total mass adsorbed as the mean of the distribution. The probability density function, f , for $a \leq x \leq b$ is given by Equations D-1 through D-3. Here, μ is the mean, σ is the standard deviation, $\phi(\xi)$ is the probability density function of the standard normal distribution, and $\Phi(\cdot)$ is the cumulative distribution function.

$$f(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right)} \quad (\text{D-1})$$

$$\phi(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right) \quad (\text{D-2})$$

$$\Phi(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(x/\sqrt{2}\right)\right) \quad (\text{D-3})$$

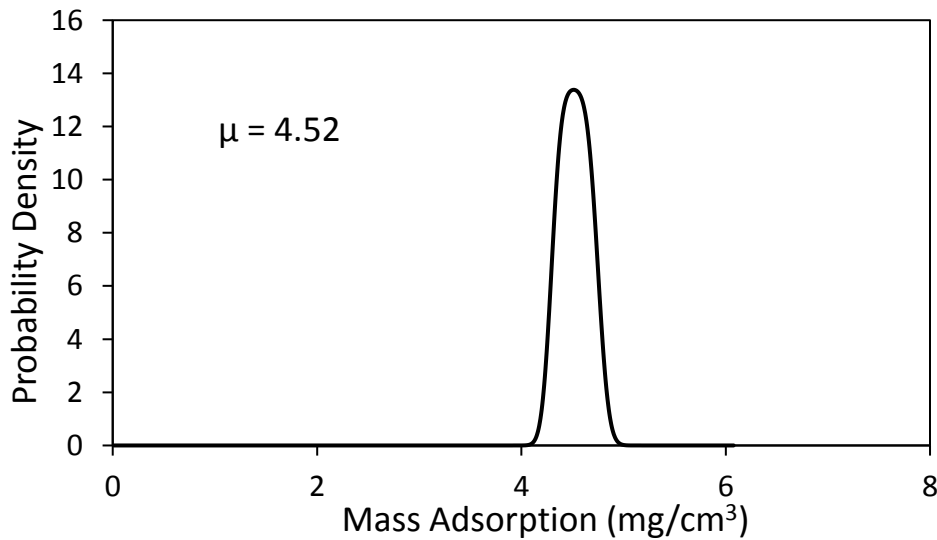


Figure 41. Truncated normal distribution for the Experiment 1 gas mixture in IRMOF-1

Consequently, we have values for probability versus total mass, which we use to assign probabilities to each gas mixture. In Figure 41 and Figure 42, the probability density is plotted against mass adsorption, having a mean equivalent to the total mass adsorbed in IRMOF-1 and ZIF-8, respectively, for the current experiment.

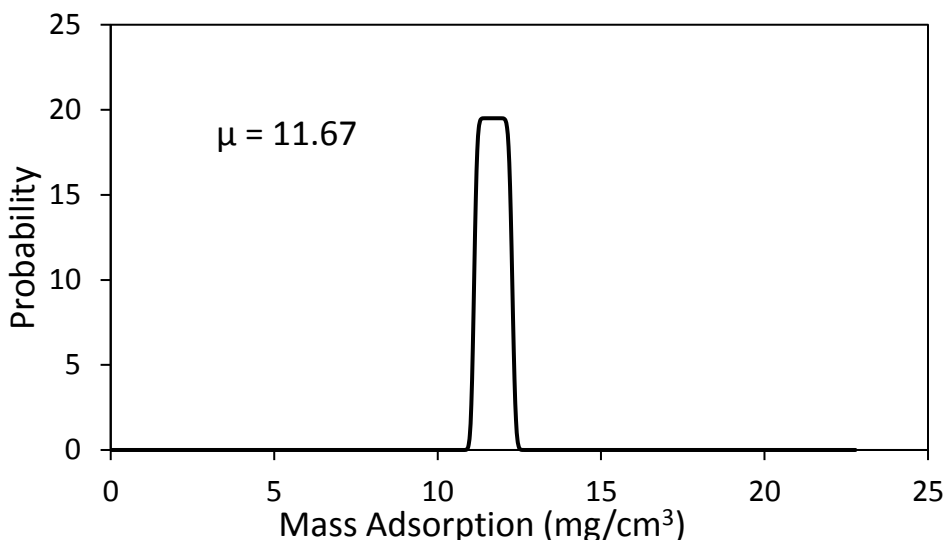


Figure 42. Truncated normal distribution for the Experiment 1 gas mixture in ZIF-8

Then, for each mixture, the simulation results for total adsorption are used to read corresponding probabilities from the above distributions, for each MOF. Each mixture then has a probability associated with it, and consequently the concentration of each gas has a probability assignment. If we plotted this data as a scatter plot it would look like Figure 43 (IRMOF-1) and Figure 44 (ZIF-8) below, for probability density versus mole fraction of CH_4 .

Next, we want to know the results if we were to have a two-MOF array of IRMOF-1 and ZIF-8. We calculate the joint probability by multiplying the probabilities at each gas mixture, when they are in discrete (probability mass function) format.

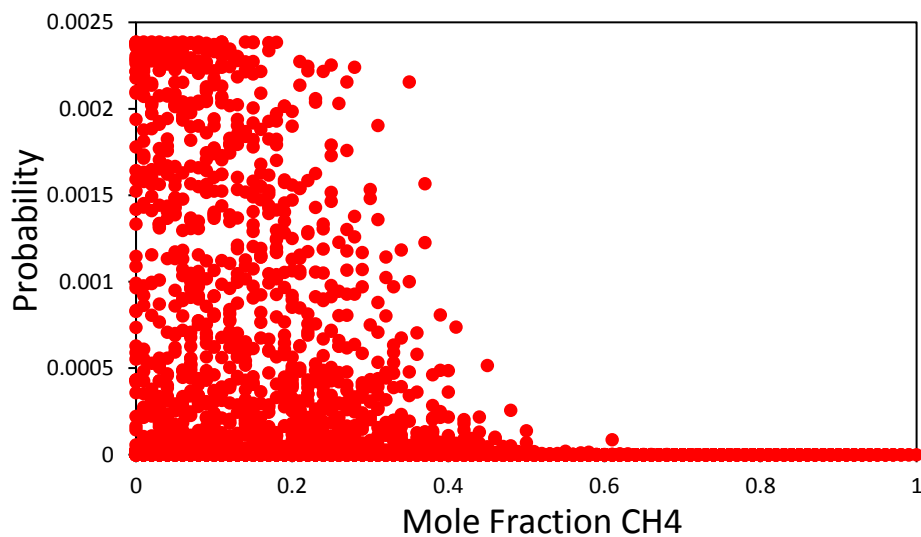


Figure 43. Probability vs mole fraction for methane in each simulated mixture in IRMOF-1

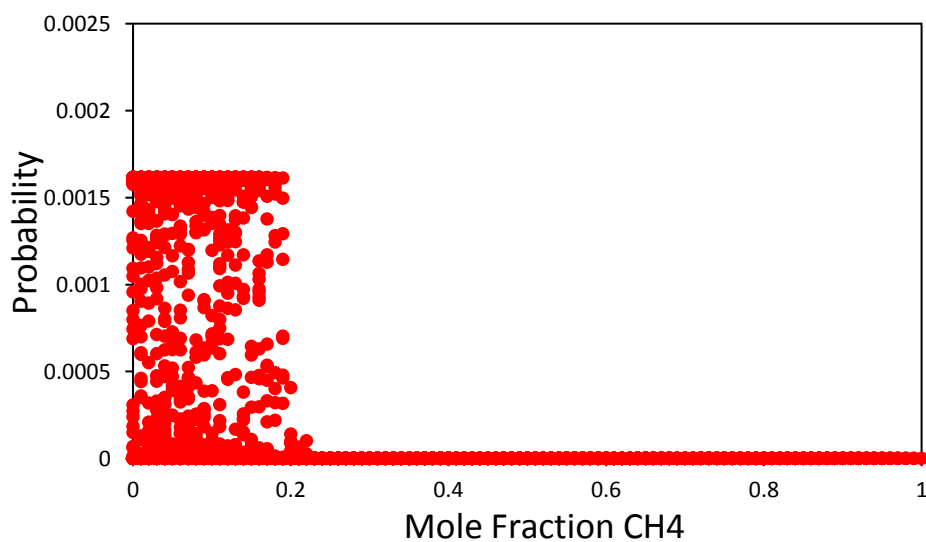


Figure 44. Probability vs mole fraction for methane in each simulated mixture in ZIF-8

Since a requirement of a set of pmfs is that their sum is 1, we normalize the probabilities after taking their product. Below is an example of this calculation.

$$P1 = [0.1, 0.2, 0.4, 0.2, 0.1]$$

$$P2 = [0.05, 0.1, 0.6, 0.15, 0.1]$$

$$\text{Joint Probability} = [0.1*0.05, 0.2*0.1, 0.4*0.6, 0.2*0.15, 0.1*0.1]$$

$$= [0.005, 0.02, 0.24, 0.03, 0.01]$$

Normalize probability so they sum to 1

$$\text{Sum} = 0.005 + 0.02 + 0.24 + 0.03 + 0.01 = 0.305$$

$$\text{Normalized probability} = \frac{0.005}{0.305} + \frac{0.02}{0.305} + \frac{0.24}{0.305} + \frac{0.03}{0.305} + \frac{0.01}{0.305}$$

$$= 0.0164 + 0.0655 + 0.787 + 0.0983 + 0.0328 = 1$$

$$\text{Final joint probability} = [0.0164, 0.0655, 0.787, 0.0983, 0.0328]$$

After the joint probability is calculated and normalized, the result is a similar scatter plot to those for the individual MOFs (Figure 43 and Figure 44), but now the probabilities are for the MOF array. It is difficult to draw any solid conclusions regarding the probable concentration of methane when the plot has this type of scattering, so we consolidate the points. For each mole fraction, we take the sum of all probabilities at that value, resulting in one probability for one mole fraction, for each gas. We then transform the probability values from discrete points to continuous

distributions, simply by multiplying each point by the total number of points (101). The following plot, Figure 45, shows the final probabilities for each gas prediction in the IRMOF-1/ZIF-8 array, after this process is done for all three gases.

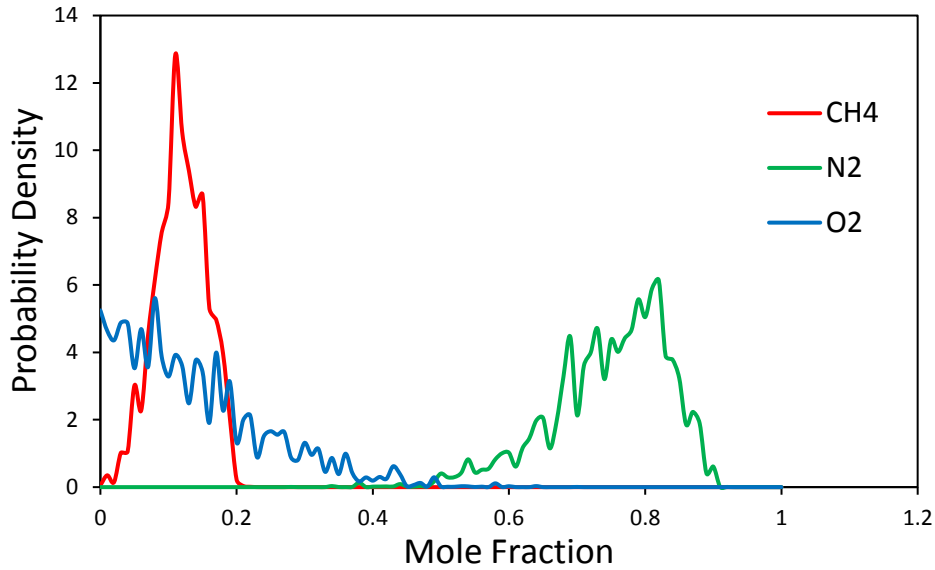


Figure 45. Probability vs Mole Fraction for CH4, O2, and N2 predicted by an array of IRMOF-1 and ZIF-8.

In our study, this calculation was performed for all combination of all 9 MOFs considered, followed by a calculation of the KLD to quantify and rank arrays. The KLD calculation for each array uses probabilities in their discrete (pmf) format. The following shows an example of the KLD calculation, using the values from the joint probability example above.

$$KLD = \sum_i^N P_i \log \frac{P_i}{Q_i}$$

$$N = 5, Q = Q_i = 1/N = 1/5 = 0.2, P = [0.0164, 0.0655, 0.787, 0.0983, 0.0328]$$

$$KLD = \left(0.0164 * \log \frac{0.0164}{0.2} + 0.0655 * \log \frac{0.0655}{0.2} + 0.787 * \log \frac{0.787}{0.2} + 0.0983 * \log \frac{0.0983}{0.2} + 0.0328 * \log \frac{0.0328}{0.2} \right)$$

$$KLD = -0.05918 + -0.10548 + 1.5554 + -0.10073 + -0.08555$$

$$KLD = 1.204$$

Therefore, the KLD for this array for predicting the unknown mixture is 1.204. If we do this KLD calculation for the individual array elements, P1 and P2 above, we get KLD values of 0.2 and 0.589, respectively, for each on its own. By combining the sensor signals, we can gather more information from the output. Similarly, for the example of IRMOF-1 and ZIF-8 provided, calculate the KLD as it relates to each gas, since they have separate probability distributions. The result is a KLD for CH₄ of 2.79, 1.68 for N₂, and 1.55 for O₂. For ranking purposes, we take the product of these three KLD values to be 7.24, the overall information content for this IRMOF-1/ZIF-8 array to predict the unknown mixture in experiment 1.

Appendix E Genetic Algorithm

Our approach for the genetic algorithm (GA) evaluation follows standard GA procedures. The detailed algorithm with example files can be found on GitHub at <https://github.com/JennaGustafson/mof-array-genetic-algorithm>. First, we will describe the overall steps of the GA, followed by details of the input specifications and the fitness function calculation.

Genetic Algorithm Procedure

1. Create first population
2. Evaluate fitness
3. Rank arrays and save top array
4. Select parents
5. For each generation
 - a. Create children
 - i. Crossover
 - ii. Mutation
 - b. Evaluate fitness
 - c. Rank arrays and save top array
 - d. Select parents

E.1 Input Files and Parameters

There are two files that are used as input when executing the algorithm, both featuring results from the aforementioned simulations. (Please note that the mass uptake from the GCMC simulations can be replaced with any sensor signal output.) The simulation outputs from RASPA

were compiled so that all results for every MOF and gas mixture were in one file (see attached examples). The other input file defines the “unknown” mixture which we are “asking” our MOF arrays to predict; moreover, this prediction ability provides the criteria we use to rank arrays (i.e. How well can a MOF array predict this unknown gas mixture, provided only a mass change from each array element?). All results presented in the manuscript are from testing a mixture of 10% methane, 20% oxygen, and 70% nitrogen.

Additionally, there are several parameters that must be defined by the user in a process configuration file, found in the Settings directory in the GitHub repository. First, we specify the *array size* we are interested in optimizing, which for this study could be anywhere from one to fifty. Then, we decide a *mutation rate*, in percent, a *population size*, and number of *generations*. These may all be adjusted at the user’s discretion, and values for this study are reported in the manuscript.

Next, we specify what is defined as the “*mrangle*”, meaning mass range, for the statistical analysis. This value is the mass range used to obtain discrete probabilities from a distribution, which requires taking the difference between two cumulative distribution functions. Essentially, the probability we calculate, for a normal distribution with mean μ , is the probability that some new value x is between $\mu - mrangle$ and $\mu + mrangle$. For this, we must also define the *standard deviation*, the value of which is up to the user and may depend on the error of the sensor device in application. In this study, we have used a *standard deviation* of 10% (of the mean) and an *mrangle* of 0.001. Further, the process configuration file contains a *list of all MOFs* to be considered in the study, as well as their *densities*. Densities are used to convert mass uptake values from mg/g to mg/cm³. This pre-processing step may be skipped depending on your needs. Finally, the *gases* in consideration are listed in the configuration file, as well.

E.2 Calculation of Fitness Function

In the following section we describe in detail the methodology for calculating the Kullback-Liebler Divergence values for arrays, which is carried out when evaluating the fitness function throughout the genetic algorithm.

For each MOF, we calculate probability distributions for their ability (as single sensors) to predict the composition of each gas in the “unknown” gas mixture. The mass value from the “unknown” mixture is read in as the signal from a MOF and is defines as the mean of a normal distribution. Below in Table 18 is an example of an “unknown” mixture and corresponding mass values, or signals, from selected MOFs.

Table 18. "Unknown" mixture 1 composition and simulation results

Component Mole Fraction			Total Mass Adsorbed (mg/cm ³)	
CH ₄	O ₂	N ₂	MOF-1	MOF-2
0.1	0.2	0.7	4.95	11.65

The next step in the calculation is to create truncated normal distributions, using each total mass as the mean of the distribution. The distribution must be truncated so that no negative mass values are possible, so the x-axis begins at 0 (0 total mass adsorption). The probability density function, f , for $a \leq x \leq b$ is given by the Equations 16-18 listed previously.

In Figure 46, the probability densities for the example MOFs are plotted against mass uptake; the mean of each distribution is the signal from the “unknown” mixture, as listed in Table 18.

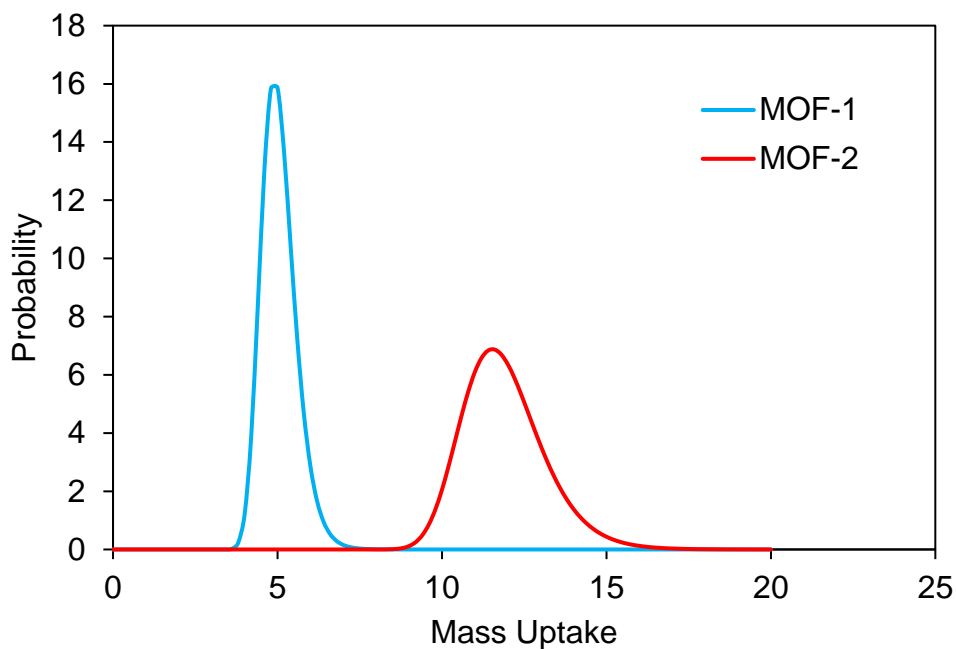


Figure 46. Normal probability distributions for example MOFs 1 and 2

The goal is to compare all of the *simulated* mass uptake results to this signal and determine how close the mass values are, defined in terms of probability. For each mixture, the simulation results for total adsorption are used to read corresponding probabilities from the above distributions, for each MOF. The calculated probabilities are discrete probabilities, also known as probability mass function (pmf) values; after all probabilities for a MOF are determined, they are normalized so that the sum over all pmf values is 1. This process results in a pmf value corresponding to each gas mixture, calculated separately for all MOFs.

Each mixture then has a probability associated with it, and consequently the concentration of each gas has a probability assignment. If we plotted this data as a scatter plot for each example MOF it would look like Figure 47, below, for probability density versus mole fraction of CH₄.

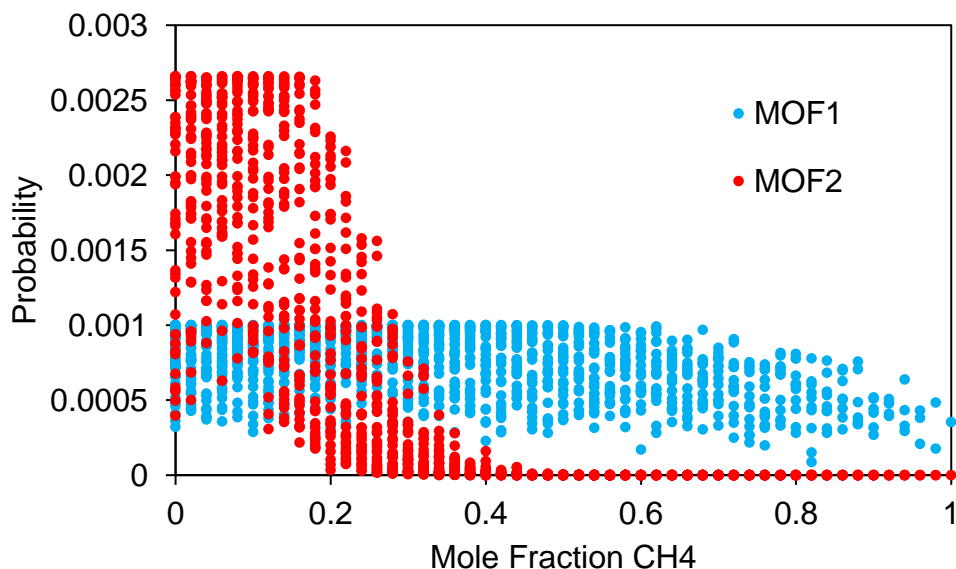


Figure 47. Probability versus concentration of methane for both example MOFs. Each point corresponds to a simulated gas mixture, for which probabilities were read from the normal distributions.

Since we have all probabilities for the individual MOFs, it is time to calculate probabilities for *arrays* of MOFs. We combine the signals from each MOF in the array using a *joint probability* calculation. The joint probability is simply taking the pmf values for all MOFs in that array and multiplying them pointwise and normalizing the result to the sum of pmfs over all concentrations is 1. Now, we have pmf vs concentration for each MOF array combination.

For example, we want to know the results if we were to have a two-MOF array. We calculate the joint probability by multiplying the probabilities at each gas mixture, when they are in discrete pmf format. Since a requirement of a set of pmfs is that their sum is 1, we normalize the probabilities after taking their product. Here is an example of this calculation:

$$P1 = [0.1, 0.2, 0.4, 0.2, 0.1]$$

$$P2 = [0.05, 0.1, 0.6, 0.15, 0.1]$$

$$\text{Joint Probability} = P1 \times P2 = [0.1*0.05, 0.2*0.1, 0.4*0.6, 0.2*0.15, 0.1*0.1]$$

$$= [0.005, 0.02, 0.24, 0.03, 0.01]$$

Normalize probability so they sum to 1

$$\text{Sum} = 0.005 + 0.02 + 0.24 + 0.03 + 0.01 = 0.305$$

$$\text{Normalized probability} = \frac{\text{Joint Probability}}{\text{Sum}} = \frac{0.005}{0.305}, \frac{0.02}{0.305}, \frac{0.24}{0.305}, \frac{0.03}{0.305}, \frac{0.01}{0.305}$$

$$\text{Final joint probability} = [0.0164, 0.0655, 0.787, 0.0983, 0.0328]$$

When we calculate the joint probability of the two example MOFs, we plot the new probabilities again versus methane concentration, shown in Figure 48.

After the joint probability is calculated and normalized, the result is a similar scatter plot to those for the individual MOFs (Figure 47), but now the probabilities are for the MOF array. It is difficult to draw any solid conclusions regarding the probable concentration of methane when the plot has this type of scattering, so we consolidate the points. When working with mixtures greater than two components, it is necessary to separate the probabilities for each gas so that we can predict the concentration of each component. Thus, we combine pmf values at each mole fraction for each gas, by taking the sum. For example, all probabilities associated with mixtures

that have a methane mole fraction of 0.2 will be added together. This is done for all gases, resulting in 51 mole fractions (0-1) for each gas, by increments of 0.02, having a pmf value for each MOF array associated with that concentration.

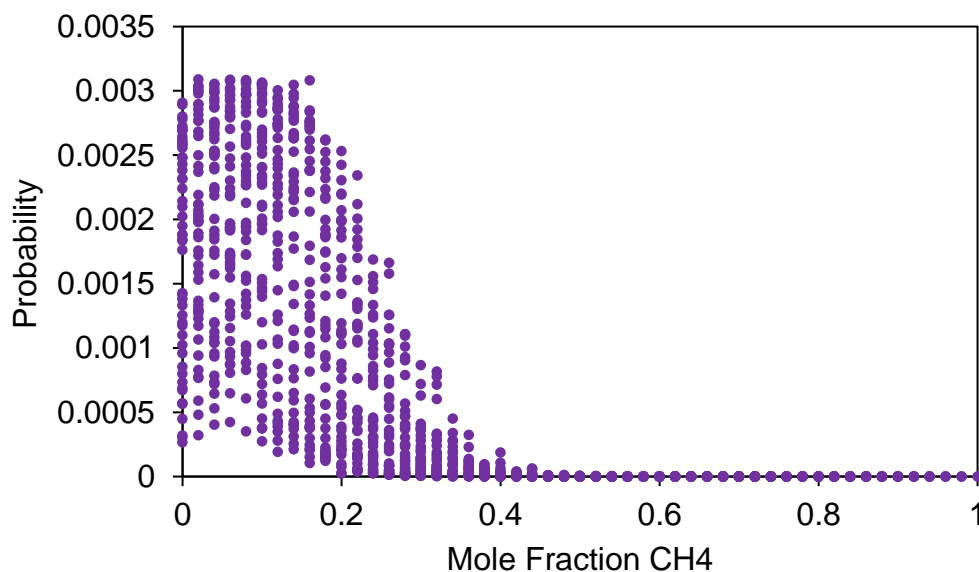


Figure 48. Probability vs methane concentration for the array of MOF-1 and MOF-2

Now that we have comprehensive probabilities for an array, we can quantify its performance for determining the “unknown” mixture. We then transform the probability values from discrete points to continuous distributions, simply by multiplying each point by the total number of points (51).

The following plot, Figure 49, shows the final probability distributions for each gas prediction in the example two MOF array, after this process is done for all three gases. Finally, for each array/gas species combination we calculate the Kullback-Liebler Divergence (KLD) value, in terms of bits of information. This information content is a relation between the pmfs (P_i)

we calculated and a reference probability ($Q_i = 1/51$). This results in three KLD values for an array's ability to predict: methane, nitrogen, and oxygen. Shown below (Figure 50) are the probability sets used to calculate the KLD.

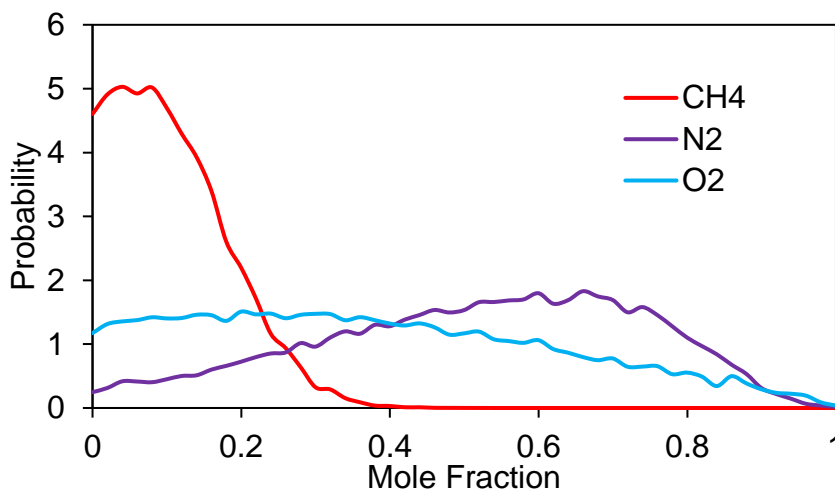


Figure 49. Probability versus concentration for CH₄, O₂, and N₂, as predicted by a two MOF array.

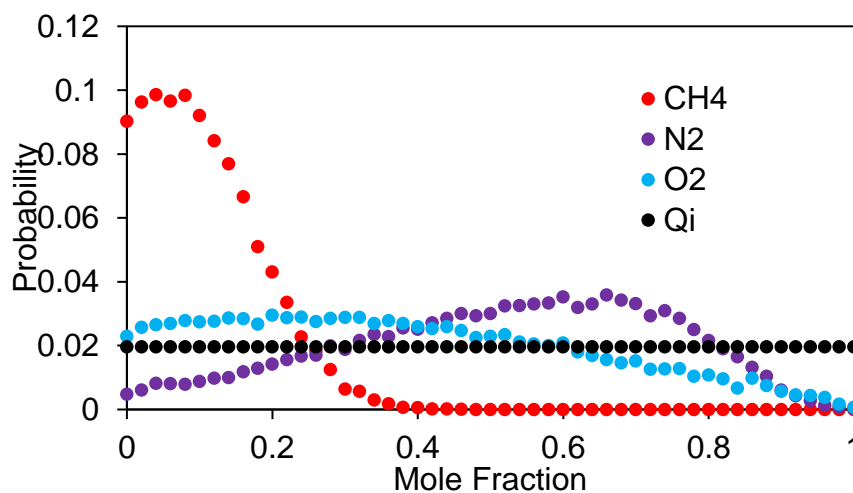


Figure 50. Probability versus concentration, including reference probability set Q (discrete version of Figure S3).

To get an overall metric for an array, we multiply the KLD values from each gas and accept this as the score for that array. The following shows a simplified example of the KLD calculation, using only five probability points. Please note we use a log of base 2 to yield units of bits,

$$KLD = \sum_i^N P_i \log \frac{P_i}{Q_i}$$

$$N = 5, Q = Q_i = 1/N = 1/5 = 0.2, P = [0.0164, 0.0655, 0.787, 0.0983, 0.0328]$$

$$KLD = \left(0.0164 * \log \frac{0.0164}{0.2} + 0.0655 * \log \frac{0.0655}{0.2} + 0.787 * \log \frac{0.787}{0.2} + 0.0983 * \log \frac{0.0983}{0.2} + 0.0328 * \log \frac{0.0328}{0.2} \right)$$

$$KLD = -0.05918 + -0.10548 + 1.5554 + -0.10073 + -0.08555$$

$$KLD = 1.204$$

This is carried out for the probability set for each gas, and then we take the product of all three KLD values. This is taken as the fitness function value for each array.

E.3 GA Analysis

Here, we provide some additional insights from the genetic algorithm results that are not critical to the outcomes featured in the manuscript.

Just as with changing the population size, we can look deeper at the effects of array size on the genetic algorithm run time. For a population size of 100 and mutation rate of 0.1%, we plot the time for a range of arrays from 10 to 25 MOFs in Figure 51.

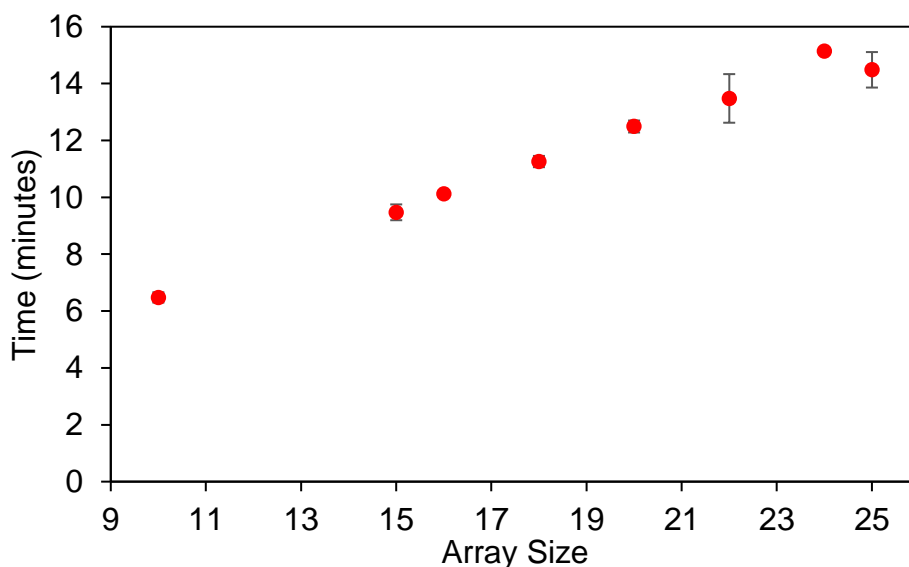


Figure 51. Plot of time in minutes versus array size for a population of 100 and mutation rate of 0.1.

The execution time steadily increases, as one would expect, where for 25 MOF arrays the algorithm takes around 15 minutes. Compared to a brute force calculation that would take years to complete or testing combinations by experimental trial-and-error, this is a trivial amount of time to determine the best candidate MOF arrays.

As previously mentioned, there are many more materials to be considered in practice besides the fifty MOFs used in this study. In Figure 52 below is a plot of algorithm execution time versus the number of MOFs from which you have to select, where the time increases linearly. This result agrees with intuition, as the larger your search space becomes, the longer it will take to search it. Therefore, a high throughput screening of MOF materials will take significantly longer,

considering that there are over 6,000 known possible MOF structures. Perhaps this will require a more efficient method with multiple screening levels, lessening the burden on the genetic algorithm portion.

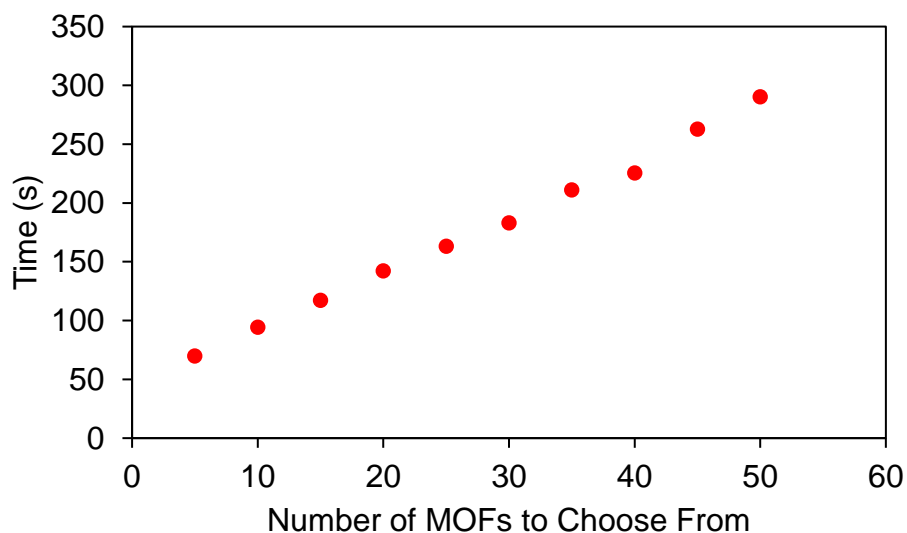


Figure 52. GA run time vs total MOFs to select from

E.4 Theoretical Limits of the Kullback-Liebler Divergence

The nature of the Kullback-Liebler divergence is that as probability distributions narrow and heighten, the KLD increases. In the analysis presented in the main manuscript, we see the KLD values increase with increasing array size. This raises the question: what is the maximum possible KLD value for this system? The best probability set would be one where there is a single mole fraction with 100% probability and the rest having essentially zero. This can be tested, where below in Figure 53 we show an example of this probability distribution. If this were the case for

all three gases in the unknown mixture of interest (i.e. 100% specific of the concentrations of each gas), that would be the maximum performance. We then calculate the KLD for this scenario, for each gas in the mixture. In this case they would all have an individual KLD around 5.67 bits. Then, we take the cube of this result to get the overall KLD for this array, which is around 182.5 bits³.

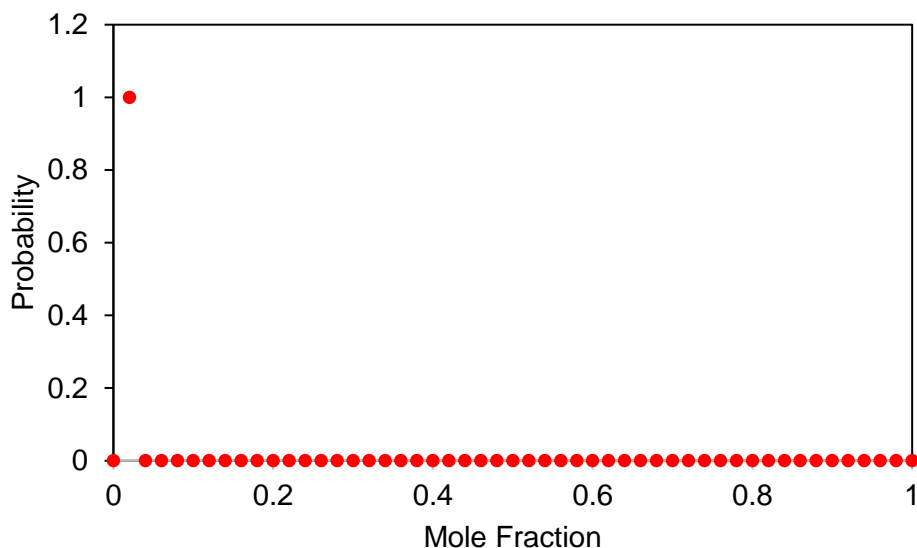


Figure 53. Probability vs mole fraction for one concentration having essentially 100% probability.

At an array size of 50 MOFs we are achieving a maximum KLD of around 45 bits³, so we are not reaching close to the maximum. Our results tell us that we should be steadily increasing in KLD as MOFs are added to the array, where the addition of MOFs beyond 50 should continue to follow this trend. Perhaps when all possible MOFs are considered, there will be a smaller array size that reaches the maximum KLD value. Please note that the maximum KLD is dependent on the step size of the mole fractions (i.e. the number of points in the probability set). For example, if we varied the gas mixtures by 1%, each individual KLD would be 6.66 bits and the overall KLD would be around 295 bits³.

Bibliography

- [1] Y. Adiguzel, H. Kulah, Breath sensors for lung cancer diagnosis, *Biosens. Bioelectron.* 65 (2015) 121–138. doi:10.1016/j.bios.2014.10.023.
- [2] K. Arshak, E. Moore, G.M. Lyons, J. Harris, S. Clifford, A review of gas sensors employed in electronic nose applications, *Sens. Rev.* 24 (2004) 181–198. doi:10.1108/02602280410525977.
- [3] W. Cao, Y. Duan, Breath Analysis: Potential for Clinical Diagnosis and Exposure Assessment, *Clin. Chem.* 52 (2006) 800–811. doi:10.1373/clinchem.2005.063545.
- [4] A.D. Wilson, M. Baietto, Applications and advances in electronic-nose technologies, *Sensors.* 9 (2009) 5099–5148. doi:10.3390/s90705099.
- [5] G. Korotcenkov, Introduction to Chemical Sensor Technologies, in: *Chem. Sens. Compr. Sens. Technol.*, Momentum Press, New York, 2011.
- [6] N. Lotfivand, M.N. Hamidon, V. Abdolzadeh, Fault tolerant architecture for artificial olfactory system, *Meas. Sci. Technol.* 26 (2015) 055101. doi:10.1088/0957-0233/26/5/055101.
- [7] C. Bur, M. Bastuck, D. Puglisi, A. Schütze, A. Lloyd Spetz, M. Andersson, Discrimination and quantification of volatile organic compounds in the ppb-range with gas sensitive SiC-FETs using multivariate statistics, *Sens. Actuators B Chem.* 214 (2015) 225–233. doi:10.1016/j.snb.2015.03.016.
- [8] B. Mumykmaz, K. Karabacak, An E-Nose-based indoor air quality monitoring system: prediction of combustible and toxic gas concentrations, *Turk. J. Electr. Eng. Comput. Sci.* 23 (2015) 729–740. doi:10.3906/elk-1304-210.
- [9] D. James, S.M. Scott, Z. Ali, W.T. O'Hare, Chemical Sensors for Electronic Nose Systems, *Microchim. Acta.* 149 (2004) 1–17. doi:10.1007/s00604-004-0291-6.
- [10] P.C. Jurs, G.A. Bakken, H.E. McClelland, Computational methods for the analysis of chemical sensor array data from volatile analytes, *Chem. Rev.* 100 (2000) 2649–2678.
- [11] K.J. Albert, N.S. Lewis, C.L. Schauer, G.A. Sotzing, S.E. Stitzel, T.P. Vaid, D.R. Walt, Cross-reactive chemical sensor arrays, *Chem. Rev.* 100 (2000) 2595–2626.
- [12] D. Compagnone, M. Faieta, D. Pizzoni, C. Di Natale, R. Paolesse, T. Van Caelenberg, B. Beheydt, P. Pittia, Quartz crystal microbalance gas sensor arrays for the quality control of chocolate, *Sens. Actuators B Chem.* 207, Part B (2015) 1114–1120. doi:10.1016/j.snb.2014.10.049.
- [13] H.P. Lang, M.K. Baller, R. Berger, C. Gerber, J.K. Gimzewski, F.M. Battiston, P. Fornaro, J.P. Ramseyer, E. Meyer, H.J. Güntherodt, An artificial nose based on a micromechanical

- cantilever array, *Anal. Chim. Acta.* 393 (1999) 59–65. doi:10.1016/S0003-2670(99)00283-4.
- [14] A. Karion, C. Sweeney, G. Pétron, G. Frost, R. Michael Hardesty, J. Kofler, B.R. Miller, T. Newberger, S. Wolter, R. Banta, A. Brewer, E. Dlugokencky, P. Lang, S.A. Montzka, R. Schnell, P. Tans, M. Trainer, R. Zamora, S. Conley, Methane emissions estimate from airborne measurements over a western United States natural gas field, *Geophys. Res. Lett.* 40 (2013) 4393–4397. doi:10.1002/grl.50811.
 - [15] G.Y. Chai, O. Lupan, E.V. Rusu, G.I. Stratan, V.V. Ursaki, V. Şontea, H. Khallaf, L. Chow, Functionalized individual ZnO microwire for natural gas detection, *Sens. Actuators Phys.* 176 (2012) 64–71. doi:10.1016/j.sna.2012.01.012.
 - [16] S.I. MacNaughton, S. Sonkusale, Gas Analysis System on Chip With Integrated Diverse Nanomaterial Sensor Array, *IEEE Sens. J.* 15 (2015) 3500–3506. doi:10.1109/JSEN.2015.2391181.
 - [17] U. Lange, N.V. Roznyatovskaya, V.M. Mirsky, Conducting polymers in chemical sensors and arrays, *Anal. Chim. Acta.* 614 (2008) 1–26. doi:10.1016/j.aca.2008.02.068.
 - [18] M.S. Freund, N.S. Lewis, A chemically diverse conducting polymer-based “electronic nose”, *Proc. Natl. Acad. Sci.* 92 (1995) 2652–2656. doi:10.1073/pnas.92.7.2652.
 - [19] A.A. Iyogun, M.R. Kumar, M.S. Freund, Chemically diverse sensor arrays based on electrochemically copolymerized pyrrole and styrene derivatives, *Sens. Actuators B Chem.* 215 (2015) 510–517. doi:10.1016/j.snb.2015.03.070.
 - [20] M.J. Kangas, R.M. Burks, J. Atwater, R.M. Lukowicz, P. Williams, A.E. Holmes, Colorimetric Sensor Arrays for the Detection and Identification of Chemical Weapons and Explosives, *Crit. Rev. Anal. Chem.* 47 (2017) 138–153. doi:10.1080/10408347.2016.1233805.
 - [21] Z. Li, J.R. Askim, K.S. Suslick, The Optoelectronic Nose: Colorimetric and Fluorometric Sensor Arrays, *Chem. Rev.* 119 (2019) 231–292. doi:10.1021/acs.chemrev.8b00226.
 - [22] P. Anzenbacher, F. Li, M.A. Palacios, Toward Wearable Sensors: Fluorescent Attoreactor Mats as Optically Encoded Cross-Reactive Sensor Arrays, *Angew. Chem. Int. Ed.* 51 (2012) 2345–2348. doi:10.1002/anie.201105629.
 - [23] A. Ranft, F. Niekietel, I. Pavlichenko, N. Stock, B.V. Lotsch, Tandem MOF-Based Photonic Crystals for Enhanced Analyte-Specific Optical Detection, *Chem. Mater.* 27 (2015) 1961–1970. doi:10.1021/cm503640c.
 - [24] H. Wohltjen, R. Dessy, Surface acoustic wave probes for chemical analysis. II. Gas chromatography detector, *Anal. Chem.* 51 (1979) 1465–1470. doi:10.1021/ac50045a025.
 - [25] J. Devkota, P.R. Ohodnicki, D.W. Greve, SAW Sensors for Chemical Vapors and Gases, *Sensors*. 17 (2017) 801. doi:10.3390/s17040801.
 - [26] R.C. AJ Ricco, Surface acoustic wave chemical sensor arrays: New chemically sensitive interfaces combined with novel cluster analysis to detect volatile organic compounds and mixtures, *Acc. Chem. Res.* 31 (1998) 289–296. doi:10.1021/ar9600749.

- [27] C.G. Fox, J.F. Alder, Surface acoustic wave sensors for atmospheric gas monitoring. A review, *Analyst*. 114 (1989) 997–1004. doi:10.1039/AN9891400997.
- [28] M.N. Hamidon, Z. Yunusa, Sensing Materials for Surface Acoustic Wave Chemical Sensors, *Progresses Chem. Sens.* (2016). doi:10.5772/63287.
- [29] M. Penza, F. Antolini, M.V. Antisari, Carbon nanotubes as SAW chemical sensors materials, *Sens. Actuators B Chem.* 100 (2004) 47–59. doi:10.1016/j.snb.2003.12.019.
- [30] L.M. Dorozhkin, I.A. Rozanov, Acoustic Wave Chemical Sensors for Gases, *J. Anal. Chem.* 56 (2001) 399–416. doi:10.1023/A:1016662616648.
- [31] W.P. Jakubik, Surface acoustic wave-based gas sensors, *Thin Solid Films*. 520 (2011) 986–993. doi:10.1016/j.tsf.2011.04.174.
- [32] M. Penza, M.A. Tagliente, L. Mirengi, C. Gerardi, C. Martucci, G. Cassano, Tungsten trioxide (WO₃) sputtered thin films for a NO_x gas sensor, *Sens. Actuators B Chem.* 50 (1998) 9–18. doi:10.1016/S0925-4005(98)00149-X.
- [33] H. Nakagawa, N. Yamamoto, S. Okazaki, T. Chinzei, S. Asakura, A room-temperature operated hydrogen leak sensor, *Sens. Actuators B Chem.* 93 (2003) 468–474. doi:10.1016/S0925-4005(03)00201-6.
- [34] K. Beck, T. Kunzelmann, M. von Schickfus, S. Hunklinger, Contactless surface acoustic wave gas sensor, *Sens. Actuators Phys.* 76 (1999) 103–106. doi:10.1016/S0924-4247(98)00359-8.
- [35] W. Wang, H. Hu, X. Liu, S. He, Y. Pan, C. Zhang, C. Dong, Development of a Room Temperature SAW Methane Gas Sensor Incorporating a Supramolecular Cryptophane A Coating, *Sensors*. 16 (2016) 73. doi:10.3390/s16010073.
- [36] L. Al-Mashat, H.D. Tran, W. Wlodarski, R.B. Kaner, K. Kalantar-zadeh, Polypyrrole nanofiber surface acoustic wave gas sensors, *Sens. Actuators B Chem.* 134 (2008) 826–831. doi:10.1016/j.snb.2008.06.030.
- [37] W.P. Jakubik, M.W. Urbańczyk, S. Kochowski, J. Bodzenta, Bilayer structure for hydrogen detection in a surface acoustic wave sensor system, *Sens. Actuators B Chem.* 82 (2002) 265–271. doi:10.1016/S0925-4005(01)01061-9.
- [38] H. Amal, D.-Y. Shi, R. Ionescu, W. Zhang, Q.-L. Hua, Y.-Y. Pan, L. Tao, H. Liu, H. Haick, Assessment of ovarian cancer conditions from exhaled breath, *Int. J. Cancer*. 136 (2015) E614–E622. doi:10.1002/ijc.29166.
- [39] E. Westenbrink, R.P. Arasaradnam, N. O’Connell, C. Bailey, C. Nwokolo, K.D. Bardhan, J.A. Covington, Development and application of a new electronic nose instrument for the detection of colorectal cancer, *Biosens. Bioelectron.* 67 (2015) 733–738. doi:10.1016/j.bios.2014.10.044.
- [40] A. Pomerantz, R. Blachman-Braun, J.A. Galnares-Olalde, R. Berebichez-Fridman, M. Capurso-García, The possibility of inventing new technologies in the detection of cancer by applying elements of the canine olfactory apparatus, *Med. Hypotheses*. 85 (2015) 160–172. doi:10.1016/j.mehy.2015.04.024.

- [41] M.E. Staymates, W.A. MacCrehan, J.L. Staymates, R.R. Kunz, T. Mendum, T.-H. Ong, G. Geurtsen, G.J. Gillen, B.A. Craven, Biomimetic Sniffing Improves the Detection Performance of a 3D Printed Nose of a Dog and a Commercial Trace Vapor Detector, *Sci. Rep.* 6 (2016) 36876. doi:10.1038/srep36876.
- [42] Y. Oh, Y. Lee, J. Heath, M. Kim, Applications of Animal Biosensors: A Review, *IEEE Sens. J.* 15 (2015) 637–645. doi:10.1109/JSEN.2014.2358261.
- [43] K. Persaud, G. Dodd, Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose, *Nature*. 299 (1982) 352. doi:10.1038/299352a0.
- [44] F. Röck, N. Barsan, U. Weimar, Electronic Nose: Current Status and Future Trends, *Chem. Rev.* 108 (2008) 705–725. doi:10.1021/cr068121q.
- [45] R. Dutta, E.L. Hines, J.W. Gardner, P. Boilot, Bacteria classification using Cyranose 320 electronic nose, *Biomed. Eng. OnLine*. 1 (2002) 4. doi:10.1186/1475-925X-1-4.
- [46] L. Spinelle, M. Gerboles, G. Kok, S. Persijn, T. Sauerwald, Review of portable and low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds, *Sensors*. 17 (2017) 1520. doi:10.3390/s17071520.
- [47] C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, G. D’Arcangelo, C. Roscioni, A. Finazzi-Agrò, A. D’Amico, Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors, *Biosens. Bioelectron.* 18 (2003) 1209–1218. doi:10.1016/S0956-5663(03)00086-1.
- [48] P. Bhattacharyya, P.K. Basu, H. Saha, S. Basu, Fast response methane sensor using nanocrystalline zinc oxide thin films derived by sol–gel method, *Sens. Actuators B Chem.* 124 (2007) 62–67. doi:10.1016/j.snb.2006.11.046.
- [49] D. Zhang, N. Yin, B. Xia, Facile fabrication of ZnO nanocrystalline-modified graphene hybrid nanocomposite toward methane gas sensing application, *J. Mater. Sci. Mater. Electron.* 26 (2015) 5937–5945. doi:10.1007/s10854-015-3165-2.
- [50] V. Galstyan, E. Comini, I. Kholmanov, G. Faglia, G. Sberveglieri, Reduced graphene oxide/ZnO nanocomposite for application in chemical gas sensors, *RSC Adv.* 6 (2016) 34225–34232. doi:10.1039/C6RA01913G.
- [51] K. Liu, L. Wang, T. Tan, G. Wang, W. Zhang, W. Chen, X. Gao, Highly sensitive detection of methane by near-infrared laser absorption spectroscopy using a compact dense-pattern multipass cell, *Sens. Actuators B Chem.* 220 (2015) 1000–1005. doi:10.1016/j.snb.2015.05.136.
- [52] J.C. Kim, H.K. Jun, J.-S. Huh, D.D. Lee, Tin oxide-based methane gas sensor promoted by alumina-supported Pd catalyst, *Sens. Actuators B Chem.* 45 (1997) 271–277. doi:10.1016/S0925-4005(97)00325-0.
- [53] Z. Wang, M. Guo, G. A. Baker, J. R. Stetter, L. Lin, A. J. Mason, X. Zeng, Methane–oxygen electrochemical coupling in an ionic liquid: a robust sensor for simultaneous quantification, *Analyst*. 139 (2014) 5140–5147. doi:10.1039/C4AN00839A.

- [54] A. Star, V. Joshi, S. Skarupo, D. Thomas, J.-C.P. Gabriel, Gas Sensor Array Based on Metal-Decorated Carbon Nanotubes, *J. Phys. Chem. B.* 110 (2006) 21014–21020. doi:10.1021/jp064371z.
- [55] H. Qi, J. Liu, J. Pionteck, P. Pötschke, E. Mäder, Carbon nanotube–cellulose composite aerogels for vapour sensing, *Sens. Actuators B Chem.* 213 (2015) 20–26. doi:10.1016/j.snb.2015.02.067.
- [56] L.E. Kreno, K. Leong, O.K. Farha, M. Allendorf, R.P. Van Duyne, J.T. Hupp, Metal–organic framework materials as chemical sensors, *Chem. Rev.* 112 (2012) 1105–1125. doi:10.1021/cr200324t.
- [57] J. Lei, R. Qian, P. Ling, L. Cui, H. Ju, Design and sensing applications of metal–organic framework composites, *TrAC Trends Anal. Chem.* 58 (2014) 71–78. doi:10.1016/j.trac.2014.02.012.
- [58] T. Zhang, S. Mubeen, N.V. Myung, M.A. Deshusses, Recent progress in carbon nanotube-based gas sensors, *Nanotechnology.* 19 (2008) 332001. doi:10.1088/0957-4484/19/33/332001.
- [59] X. Xu, J. Wang, Y. Long, Zeolite-based Materials for Gas Sensors, *Sensors.* 6 (2006) 1751–1764. doi:10.3390/s6121751.
- [60] H.S. Kim, Zeolite Y film as a versatile material for electrochemical sensors, *Mater. Lett.* 182 (2016) 201–205. doi:10.1016/j.matlet.2016.06.114.
- [61] K. Sahner, G. Hagen, D. Schönauer, S. Reiß, R. Moos, Zeolites — Versatile materials for gas sensors, *Solid State Ion.* 179 (2008) 2416–2423. doi:10.1016/j.ssi.2008.08.012.
- [62] P. Kumar, A. Deep, K.-H. Kim, Metal organic frameworks for sensing applications, *TrAC Trends Anal. Chem.* 73 (2015) 39–53. doi:10.1016/j.trac.2015.04.009.
- [63] M.G. Campbell, M. Dinca, Metal-Organic Frameworks as Active Materials in Electronic Sensor Devices, *Sensors.* 17 (2017) 1108. doi:10.3390/s17051108.
- [64] J. Hromadka, B. Tokay, R. Correia, S.P. Morgan, S. Korposh, Highly sensitive volatile organic compounds vapour measurements using a long period grating optical fibre sensor coated with metal organic framework ZIF-8, *Sens. Actuators B Chem.* 260 (2018) 685–692. doi:10.1016/j.snb.2018.01.015.
- [65] W. P. Lustig, S. Mukherjee, N. D. Rudd, A. V. Desai, J. Li, S. K. Ghosh, Metal–organic frameworks: functional luminescent and photonic materials for sensing applications, *Chem. Soc. Rev.* 46 (2017) 3242–3285. doi:10.1039/C6CS00930A.
- [66] S. Achmann, G. Hagen, J. Kita, I.M. Malkowsky, C. Kiener, R. Moos, Metal-Organic Frameworks for Sensing Applications in the Gas Phase, *Sensors.* 9 (2009) 1574–1589. doi:10.3390/s90301574.
- [67] J. Devkota, K.-J. Kim, P.R. Ohodnicki, J.T. Culp, D.W. Greve, J.W. Lekse, Zeolitic imidazole framework-coated acoustic sensors for room temperature detection of carbon dioxide and methane, *ArXiv171208468 Phys.* (2017). <http://arxiv.org/abs/1712.08468>.

- [68] T.R. Zeitler, M.D. Allendorf, J.A. Greathouse, Grand Canonical Monte Carlo Simulation of Low-Pressure Methane Adsorption in Nanoporous Framework Materials for Sensing Applications, *J. Phys. Chem. C*. 116 (2012) 3492–3502. doi:10.1021/jp208596e.
- [69] I. Ellern, A. Venkatasubramanian, J.-H. Lee, P. Hesketh, V. Stavila, A. Robinson, M. Allendorf, HKUST-1 coated piezoresistive microcantilever array for volatile organic compound sensing, *Micro Amp Nano Lett.* 8 (2013) 766–769. doi:10.1049/mnl.2013.0390.
- [70] M.G. Campbell, S.F. Liu, T.M. Swager, M. Dincă, Chemiresistive Sensor Arrays from Conductive 2D Metal–Organic Frameworks, *J. Am. Chem. Soc.* 137 (2015) 13780–13783. doi:10.1021/jacs.5b09600.
- [71] M.-S. Yao, W.-X. Tang, G.-E. Wang, B. Nath, G. Xu, MOF thin film-coated metal oxide nanowire array: Significantly improved chemiresistor sensor performance, *Adv. Mater.* 28 (2016) 5229–5234. doi:10.1002/adma.201506457.
- [72] Y. Peng, V. Krungleviciute, I. Eryazici, J.T. Hupp, O.K. Farha, T. Yildirim, Methane Storage in Metal–Organic Frameworks: Current Records, Surprise Findings, and Challenges, *J. Am. Chem. Soc.* 135 (2013) 11887–11894. doi:10.1021/ja4045289.
- [73] W. Zhou, H. Wu, M.R. Hartman, T. Yildirim, Hydrogen and Methane Adsorption in Metal–Organic Frameworks: A High-Pressure Volumetric Study, *J. Phys. Chem. C*. 111 (2007) 16131–16137. doi:10.1021/jp074889i.
- [74] D. Saha, S. Deng, Hydrogen adsorption on metal-organic framework MOF-177, *Tsinghua Sci. Technol.* 15 (2010) 363–376. doi:10.1016/S1007-0214(10)70075-4.
- [75] T.R. Zeitler, T.V. Heest, D.S. Sholl, M.D. Allendorf, J.A. Greathouse, Predicting Low-Pressure O₂ Adsorption in Nanoporous Framework Materials for Sensing Applications, *ChemPhysChem*. 14 (2013) 3740–3750. doi:10.1002/cphc.201300682.
- [76] A. Martín-Calvo, E. García-Pérez, J.M. Castillo, S. Calero, Molecular simulations for adsorption and separation of natural gas in IRMOF-1 and Cu-BTC metal-organic frameworks, *Phys. Chem. Chem. Phys.* 10 (2008) 7085–7091. doi:10.1039/B807470D.
- [77] J.A. Greathouse, N.W. Ockwig, L.J. Criscenti, T.R. Guilinger, P. Pohl, M.D. Allendorf, Computational screening of metal–organic frameworks for large-molecule chemical sensing, *Phys. Chem. Chem. Phys.* 12 (2010) 12621–12629. doi:10.1039/C0CP00092B.
- [78] W. Yang, P. Wan, M. Jia, J. Hu, Y. Guan, L. Feng, A novel electronic nose based on porous In₂O₃ microtubes sensor array for the discrimination of VOCs, *Biosens. Bioelectron.* 64 (2015) 547–553. doi:10.1016/j.bios.2014.09.081.
- [79] M. García, M.J. Fernández, J.L. Fontecha, J. Lozano, J.P. Santos, M. Aleixandre, I. Sayago, J. Gutiérrez, M.C. Horrillo, Differentiation of red wines using an electronic nose based on surface acoustic wave devices, *Talanta*. 68 (2006) 1162–1165. doi:10.1016/j.talanta.2005.07.031.
- [80] G.C. Osbourn, J.W. Bartholomew, A.J. Ricco, G.C. Frye, Visual-Empirical Region-of-Influence Pattern Recognition Applied to Chemical Microsensor Array Selection and Chemical Analysis, *Acc. Chem. Res.* 31 (1998) 297–305. doi:10.1021/ar970070j.

- [81] R. Haddad, H. Lapid, D. Harel, N. Sobel, Measuring smells, *Curr. Opin. Neurobiol.* 18 (2008) 438–444. doi:10.1016/j.conb.2008.09.007.
- [82] T.K. Alkasab, J. White, J.S. Kauer, A computational system for simulating and analyzing arrays of biological and artificial chemical sensors, *Chem. Senses.* 27 (2002) 261–275. doi:10.1093/chemse/27.3.261.
- [83] Hailian Li, Eddaoudi, Design and synthesis of an exceptionally stable and highly porous metal-organic framework, *Nature.* 402 (1999) 276.
- [84] S.S.Y. Chui, S.M.F. Lo, J.P.H. Charmant, A.G. Orpen, I.D. Williams, A chemically functionalizable nanoporous material [Cu-3(TMA)(2)(H2O)(3)](n), *Science.* 283 (1999) 1148–1150. doi:10.1126/science.283.5405.1148.
- [85] C.E. Wilmer, O.K. Farha, T. Yildirim, I. Eryazici, V. Krungleviciute, A.A. Sarjeant, R.Q. Snurr, J.T. Hupp, Gram-scale, high-yield synthesis of a robust metal–organic framework for storing methane and other gases, *Energy Environ. Sci.* 6 (2013) 1158–1163. doi:10.1039/C3EE24506C.
- [86] J.H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, K.P. Lillerud, A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability, *J. Am. Chem. Soc.* 130 (2008) 13850–13851. doi:10.1021/ja8057953.
- [87] K.S. Park, Z. Ni, A.P. Côté, J.Y. Choi, R. Huang, F.J. Uribe-Romo, H.K. Chae, M. O’Keeffe, O.M. Yaghi, Exceptional chemical and thermal stability of zeolitic imidazolate frameworks, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 10186–10191.
- [88] M.G. Martin, J.I. Siepmann, Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes, *J. Phys. Chem. B.* 102 (1998) 2569–2577. doi:10.1021/jp972543+.
- [89] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *J. Am. Chem. Soc.* 114 (1992) 10024–10035. doi:10.1021/ja00051a040.
- [90] J.J. Potoff, J.I. Siepmann, Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen, *AIChE J.* 47 (2001) 1676–1682. doi:10.1002/aic.690470719.
- [91] S.L. Mayo, B.D. Olafson, W.A. Goddard, DREIDING: a generic force field for molecular simulations, *J. Phys. Chem.* 94 (1990) 8897–8909. doi:10.1021/j100389a010.
- [92] J.R. Karra, K.S. Walton, Molecular Simulations and Experimental Studies of CO₂, CO, and N₂ Adsorption in Metal–Organic Frameworks, *J. Phys. Chem. C.* 114 (2010) 15735–15740. doi:10.1021/jp105519h.
- [93] J.A. Mason, M. Veenstra, J.R. Long, Evaluating metal–organic frameworks for natural gas storage, *Chem. Sci.* 5 (2013) 32–51. doi:10.1039/C3SC52633J.
- [94] H.V. Shurmer, J.W. Gardner, Odour discrimination with an electronic nose, *Sens. Actuators B Chem.* 8 (1992) 1–11. doi:10.1016/0925-4005(92)85001-D.

- [95] J.A. Gustafson, C.E. Wilmer, Computational design of metal–organic framework arrays for gas sensing: Influence of array size and composition on sensor performance, *J. Phys. Chem. C*. 121 (2017) 6033–6038. doi:10.1021/acs.jpcc.6b09740.
- [96] D. Dubbeldam, S. Calero, D.E. Ellis, R.Q. Snurr, RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials, *Mol. Simul.* 42 (2016) 81–101. doi:10.1080/08927022.2015.1010082.
- [97] S.R. Caskey, A.G. Wong-Foy, A.J. Matzger, Dramatic tuning of carbon dioxide uptake via metal substitution in a coordination polymer with cylindrical pores, *J. Am. Chem. Soc.* 130 (2008) 10870–10871. doi:10.1021/ja8036096.
- [98] O.K. Farha, A.Ö. Yazaydin, I. Eryazici, C.D. Malliakas, B.G. Hauser, M.G. Kanatzidis, S.T. Nguyen, R.Q. Snurr, J.T. Hupp, De novo synthesis of a metal–organic framework material featuring ultrahigh surface area and gas storage capacities, *Nat. Chem.* 2 (2010) 944–948. doi:10.1038/nchem.834.
- [99] H.K. Chae, D.Y. Siberio-Pérez, J. Kim, Y. Go, M. Eddaoudi, A.J. Matzger, M. O’Keeffe, O.M. Yaghi, A route to high surface area, porosity and inclusion of large molecules in crystals, *Nature*. 427 (2004) 523–527. doi:10.1038/nature02311.
- [100] H. Furukawa, F. Gándara, Y.-B. Zhang, J. Jiang, W.L. Queen, M.R. Hudson, O.M. Yaghi, Water adsorption in porous metal–organic frameworks and related materials, *J. Am. Chem. Soc.* 136 (2014) 4369–4381. doi:10.1021/ja500330a.
- [101] S.P. Collins, T.D. Daff, S.S. Piotrkowski, T.K. Woo, Materials design by evolutionary optimization of functional groups in metal-organic frameworks, *Sci. Adv.* 2 (2016) e1600954. doi:10.1126/sciadv.1600954.
- [102] T.C. Le, D.A. Winkler, Discovery and Optimization of Materials Using Evolutionary Approaches, *Chem. Rev.* 116 (2016) 6107–6132. doi:10.1021/acs.chemrev.5b00691.
- [103] T. Dantzig, *Number: The Language of Science*, Penguin, 2007.
- [104] S. Khuri, T. Bäck, J. Heitkötter, The Zero/One Multiple Knapsack Problem and Genetic Algorithms, in: *Proc. 1994 ACM Symp. Appl. Comput.*, ACM, New York, NY, USA, 1994: pp. 188–193. doi:10.1145/326619.326694.
- [105] P.C. Chu, J.E. Beasley, A Genetic Algorithm for the Multidimensional Knapsack Problem, *J. Heuristics*. 4 (1998) 63–86. doi:10.1023/A:1009642405419.
- [106] G.R. Raidl, An improved genetic algorithm for the multiconstrained 0-1 knapsack problem, in: *1998 IEEE Int. Conf. Evol. Comput. Proc. IEEE World Congr. Comput. Intell. Cat No98TH8360*, 1998: pp. 207–211. doi:10.1109/ICEC.1998.699502.
- [107] Y.G. Chung, J. Camp, M. Haranczyk, B.J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O.K. Farha, D.S. Sholl, R.Q. Snurr, Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals, *Chem. Mater.* 26 (2014) 6185–6192. doi:10.1021/cm502594j.

- [108] J.A. Gustafson, C.E. Wilmer, Optimizing information content in MOF sensor arrays for analyzing methane-air mixtures, *Sens. Actuators B Chem.* 267 (2018) 483–493. doi:10.1016/j.snb.2018.04.049.
- [109] D. Whitley, An overview of evolutionary algorithms: practical issues and common pitfalls, *Inf. Softw. Technol.* 43 (2001) 817–831. doi:10.1016/S0950-5849(01)00188-4.
- [110] H. Guerin, H. Le Poche, R. Pohle, E. Buitrago, M. Fernández-Bolaños Badía, J. Dijon, A.M. Ionescu, Carbon nanotube gas sensor array for multiplex analyte discrimination, *Sens. Actuators B Chem.* 207, Part A (2015) 833–842. doi:10.1016/j.snb.2014.10.117.
- [111] Y. Lu, J. Li, J. Han, H.-T. Ng, C. Binder, C. Partridge, M. Meyyappan, Room temperature methane detection using palladium loaded single-walled carbon nanotube sensors, *Chem. Phys. Lett.* 391 (2004) 344–348. doi:10.1016/j.cplett.2004.05.029.
- [112] Stavila Vitalie, Schneider Christian, Mowry Curtis, Zeitler Todd R., Greathouse Jeffery A., Robinson Alex L., Denning Julie M., Volponi Joanne, Leong Kirsty, Quan William, Tu Min, Fischer Roland A., Allendorf Mark D., *Sensors: Thin Film Growth of nbo MOFs and their Integration with Electroacoustic Devices (Adv. Funct. Mater. 11/2016)*, *Adv. Funct. Mater.* 26 (2016) 1669–1669. doi:10.1002/adfm.201670066.
- [113] H. Wohltjen, Mechanism of operation and design considerations for surface acoustic wave device vapour sensors, *Sens. Actuators.* 5 (1984) 307–325. doi:10.1016/0250-6874(84)85014-3.
- [114] L. Zhang, J.I. Siepmann, Direct calculation of Henry's law constants from Gibbs ensemble Monte Carlo simulations: nitrogen, oxygen, carbon dioxide and methane in ethanol, *Theor. Chem. Acc.* 115 (2006) 391–397. doi:10.1007/s00214-005-0073-1.
- [115] C.E. Wilmer, K.C. Kim, R.Q. Snurr, An extended charge equilibration method, *J. Phys. Chem. Lett.* 3 (2012) 2506–2511. doi:10.1021/jz3008485.
- [116] Nitrogen, (n.d.). <http://webbook.nist.gov/cgi/cbook.cgi?ID=C7727379&Mask=1000> (accessed March 19, 2016).